Genetic structure and diversity of *Sorghum bicolor* at three geographical scales in Africa

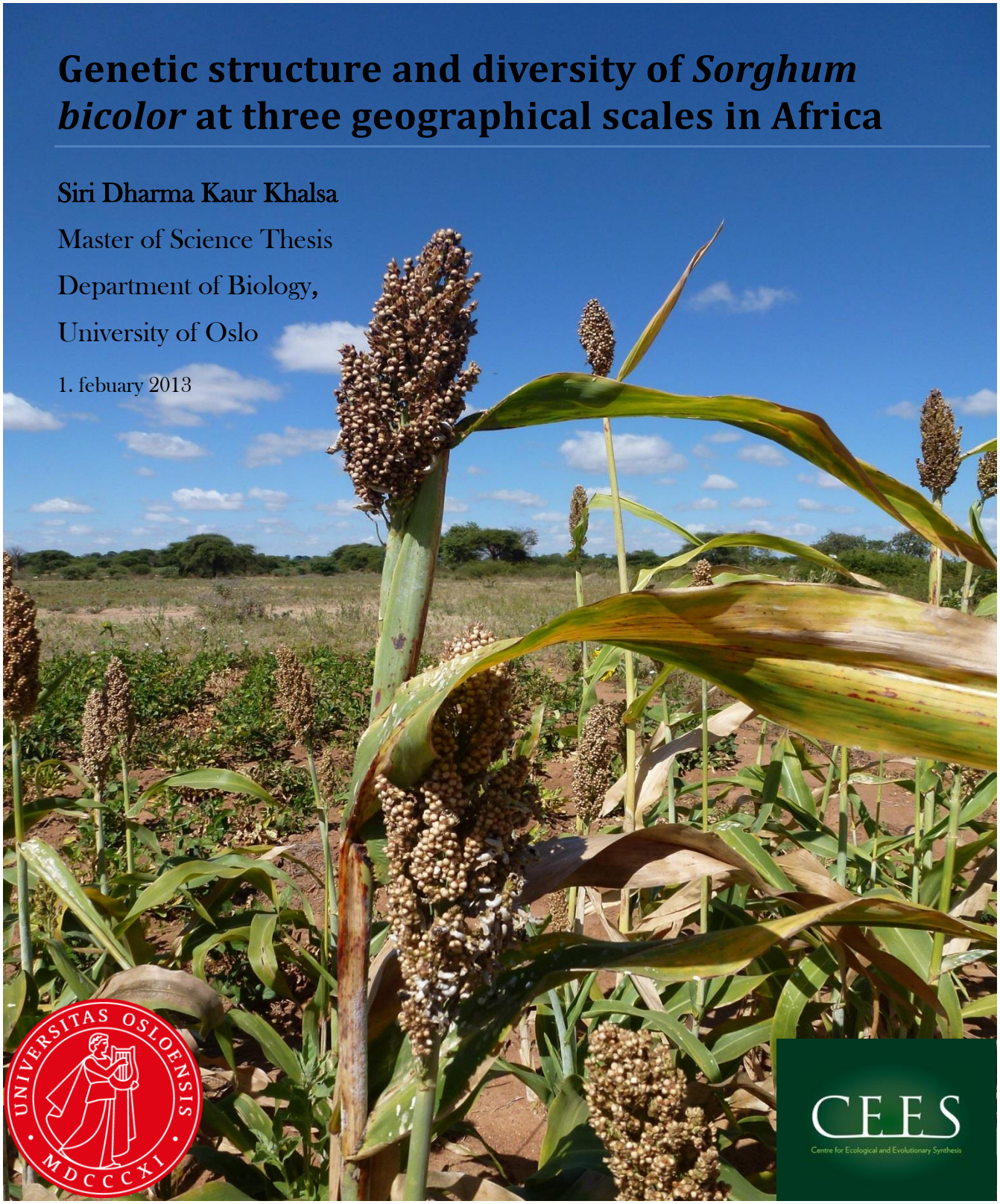
Siri Dharma Kaur Khalsa

Master of Science Thesis

Department of Biology,

University of Oslo

1. febuary 2013



Acknowledgements

Tusen takk til Anne og Ola for kjempe veiledning! Anne- du er den flotteste veilederen som finnes! Jeg vil spesielt takke for hyggelige møter, opplæring av lab metoder, støtten din under vurderinger av lab feil, hjelp med å få structure til å fungere, kaker og te og mange runder med kommentarer og glimrende forslag, som var uvurderlig i formingen av oppgaven, spesielt i slutt fasen!!! Og en stor takk til deg Ola for den sentrale rollen du har spilt for skapingen av oppgaven og felt arbeidet i Tanzania!! Takk for hyggelig isolerings tider, bestilling av frø, henting av klima data, hjelp med analyser (spesielt distruct), laging av kart, de gode kommentarene på oppgaven mot slutten og den positive instillingen din hele veien igjennom. Stor takk til Unni for hjelp med mikrosattelliter og R analyser. Takk Nanna for kjøring av fragment analyser på platene mine.

Takk til Fang, Nita og Mariel for en flott studie tid sammen!

Takk til alle de flotte menneskene på lesesal-plassen som har bidratt til et hyggelig arbeidsmiljø- Siri (morsomt at du har vært her på slutt fasen og), Ane (det var et flott initiative med vaffel møter; sånne tider kommer jeg sent til å glemme), Idunn og de nye innspillerne Marie og Mildrid- og Takk til resten av plante gruppen for mange hyggelige tirsdags møter.

Thank you to my lovely family for being the amazing family that you are and for economic, and most importantly, moral and emotional support. Thank you Guru simrat and Guru amrit for looking over the thesis at the end.

Table of Contents

ABSTRACT

Sorghum is ranked the fifth most produced food crop in the world, and is a dietary staple for over 500 million people in over 30 countries. It is the second most produced food crop in Africa, where cultivation of local varieties (landraces) of sorghum is the predominant form of agriculture.

This study investigated the genetic diversity and structure of 161 sorghum accessions, which included landraces and wild/weedy sorghum, using 17 microsatellites. The material represented three geographical scales. For the continental and country scale studies, landrace accessions from throughout Africa and throughout Tanzania were obtained from gene banks (ICRISAT and NPGRC). For the local scale study, eight landraces and wild/weedy sorghum were collected from five households in Hombolo, Tanzania.

The genetic diversity of sorghum at all three geographical scales was found to be mainly structured according to geography and less structured according to race, temperature and precipitation. At the continental scale, the accessions were (based on STRUCTURE analysis) largely divided into an eastern, western, northeastern and southern group. However, accessions from Sudan were found in all four groups, a result which supports the suggested origin of domesticated sorghum in northeastern Africa. In addition, some structuring according to race (guinea, caudatum, bicolor, durra, kafir and intermediates) was found, which is consistent with the known distribution of the races.

The cultivated accessions from Hombolo were genetically structured according to landrace and for the most part differently named landraces were genetically distinct. In addition, there was some geographical structuring of genetic diversity for the cultivated accessions (even though fields were only from 150m to 1.6 km apart), but not for wild/weedy sorghum growing alongside the sorghum crop fields. This could be explained by higher outcrossing rates in wild/weedy sorghum compared with cultivated sorghum. Considerable gene flow was detected between wild/weedy and landraces based on genetic overlap, no significant differences in genetic diversity and the number of migrating individuals. Gene flow was higher between cultivated and wild/weedy sorghum than between different landraces of cultivated sorghum.

The mainly geographical structuring of sorghum diversity can be explained by traditional cultivation practices based on indigenous landraces and a self-fertilizing mating strategy. This reflects the wealth of diversity found amongst indigenous landraces, which is important to conserve for present and future food security needs.

Keywords: sorghum, landrace, genetic diversity, wild/weedy sorghum, microsatellites.

INTRODUCTION

Climate change is expected to affect agriculture worldwide, although studies have shown that the effects will not be uniform across the globe. Small consequences are expected for crop production in developed countries with high-input agriculture, whereas for developing countries climate change is projected to have more severe effects (Lobell, et al. 2008; Parry, et al. 2007; Rosenzweig and Parry 1994). The majority of the population in Africa are subsistence farmers, and rely on rainfed low-input agriculture for their food and livelihood. Surface temperatures on the Indian ocean have risen dramatically since 1980. These changes have been shown to be correlated with droughts and growing season rainfall reductions in food insecure areas of Africa (Funk, et al. 2008). The semi-arid regions of Africa, where half the population is deemed extremely poor (UNDP 2012), are projected to see water stress and yield reductions for many important crops in the next couple of decades (Funk, et al. 2008; Knox, et al. 2012; Schlenker and Lobell 2010). Research has suggested that the best way to mitigate climate impacts for agrarians is by investments in agriculture (Funk, et al. 2008). Sorghum, pearl millet, maize and cassava are the most important rainfed cereals/root vegetables of Africa. Other important, but not as widely distributed staples are barley, tef, fonio and finger millet (Murdock 1960). Sorghum is the dietary staple for 500 million people in over 30 countries (ICRISAT 2010), and is especially important in semi-arid regions of Africa, along with pearl millet and cassava. Sorghum, pearl millet and cassava are all grown where rainfall is insufficient to support maize crops. Sorghum and pearl millet compare favorably to other grains under high-input agriculture, while they are superior to other grains under low-input agriculture (FAO 1999). The importance of sorghum, as well as pearl millet and cassava (as they are tolerant to heat and drought) is projected to increase as temperatures rise, or water becomes a limiting resource, and suitable areas for the cultivation of favorable crops such as wheat, rice, barley and maize are reduced. When comparing projected future temperature shifts for different countries in Africa, it was found that current temperatures in Tanzania, Sudan, Cameroon, Kenya and Nigeria were analogous to those projected for many other areas in Africa over the next couple decades (Burke, et al. 2009) thus, it is important to do research on crops in semi-arid regions, as they may be an important resource for other areas as temperatures rise.

Landraces and improved varieties

Before the 1950s cereal cultivation was comprised of local farmer's varieties called landraces (Duncan, et al. 1995). The term landrace is either used for a cultivated plant which has an obscure origin, is locally/environmentally adapted, or has not been through a formal breeding program (Berg 2009). The rediscovery of Mendel's laws of genetics in the beginning of the last century, opened up a new era of crop breeding. This resulted in the release of high yielding crop varieties produced by professional plant breeders (from now on 'improved varieties') through the exploitation of hybrid vigor (Ball 1930; Evenson and Gollin 2003; Swaminathan 2006). The success of these improved varieties became characterized as the 'green revolution'. Sorghum improvement in the 1950-1960s doubled yields of sorghum in India and China, and quadrupled yields in the United States (Evenson and Gollin 2003; Li and Li 1998; Vietmeyer 1996). The success of China's sorghum production increase is largely attributed to the use of landraces (which are environmentally adapted) for the development of improved varieties. In 1951 an intensive selection of landraces was conducted among villages. This resulted in the release of improved varieties that immediately increased sorghum yields by 10%, which steadily increased as new varieties were introduced (Li and Li 1998).

Sorghums yield potential exceeds that of rice wheat and maize. Given the right conditions yields have been recorded up to 13000 kg/ha, with standard yield under high-input conditions being between 3000-9000 kg/ha (House 1985). Despite large areas of land devoted to sorghum production in Africa yields of sorghum are low, averaging 700 kg/ha (UNDP 2012), compared to countries which use improved varieties and high-input agriculture. Africa has not experienced the green revolution the way China and India has (Botha and Viljoen 2008), and cultivation of indigenous landraces has remained the dominant form of agriculture. Farmers in Africa rely on a wide variety of landraces to cope with climate, diseases, pests and soil limitations, in the absence of pesticides, inputs and improved varieties (Cavatassi, et al. 2011; Vigouroux, et al. 2011). The agriculture conditions in Africa can be attributed to low governmental investment in agriculture, poverty and the yield instability of introduced improved varieties. African improved varieties have not been reliable under low-input agriculture and variable environmental conditions, and have therefore not been adopted by farmers to any large degree (Ahmed, et al. 2000; Evenson and Gollin 2003; FAO 1999; Seboka and van Hintum 2006).

In addition to low yield sorghum production is also affected by biotic and abiotic factors such as, diseases (downy mildew), pests (quelea, striga, shoot fly, sorghum midge, and sugarcane aphid), drought and soil infertility/acidity. Studies have found resistance to different biotic and abiotic factors amongst different landraces and wild/weedy sorghums (Kamala, et al. 2009; Kamala, et al. 2002; Maqbool, et al. 2001; Rai, et al. 1999; Rich, et al. 2004; Ringo 2009; Vietmeyer 1996). It is believed that sorghum yields and/or further resistance to biotic and abiotic factors could be improved upon, given the availability of suitable improved varieties, and preferably developed from environmentally suited local landraces and/or wild/weedy varieties (Makanda, et al. 2010). It has also been suggested that in some cases aiding farmers in cultivation techniques and equipment is more beneficial than introducing improved varieties (Vietmeyer 1996). Thus, understanding the motivations and constraints of farmers in cultivation practices and adoption of improved varieties is necessary in order to effectively improve sorghum production in Africa (Cavatassi, et al. 2011).

Origin and domestication

Sorghum, pearl millet and finger millet have their center of diversity in Africa, and were all domesticated from African wild progenitors (Brunken, et al. 1977; Doggett 1991; Mehra 1991). Archeological evidence for wild sorghum has been found at four excavation sites; grains were found from 105,000 years ago in Mozambique at the Ngalue cave site (Mercader 2009), 8000-9000 years ago in southern Egypt at the Nabta playa site (Wasylikowa and Dahlberg 1999), 7000 years ago in the Nile valley at the Farafra site and 6000 years ago in Egypt at the Abu Ballas site (Barakat and Fahmy 1999). The first traces of cultivated sorghums grains were found in China, the oldest dating back to 7000 BP (Kimber 2000). The oldest cultivated sorghum found in Africa are impressions of spikelet's from sorghum, together with pearl millet and finger millet, on pots in Kadero, Sudan dating back to around 6000 BP (Klichowska 1984). Cultivated sorghum was found in India around 4000 BP (Kajale 1977). The first evidence of cultivated sorghum grains in Africa were found in eastern Sudan dating back to around 3000 BP (Fattovich, et al. 1984) and central Sudan dating back to around 18000 ± 1400 BP (Clark and Stemler 1975). Despite the archeological findings, sorghum is thought to have been domesticated in Africa before it was transported to China and India (Li, et al. 2010), owing to the fact that cultivated sorghum arose from wild sorghum, which has been confined to Africa up until recent times (De Wet and Harlan 1971; Doggett 1991). The exact time and place of sorghum domestication is still unsubstantiated,

but most are of the opinion that it was domesticated somewhere in the northeast African area. It has been claimed that it was first domesticated in western Sudan (Murdock 1960), central Sudan (De Wet and Harlan 1971) and Ethiopia (Doggett 1991).

Sorghum description

Sorghum is a C4 annual wind pollinated cereal with high photosynthetic efficiency. It varies in form and can be from 50 cm to 6 m tall, usually with a large erect stem terminating in a head or panicle with variable compactness. The leaves look similar in appearance to maize and during drought they curl inward as to reduce moisture loss through transpiration. The root system is deeply penetrating, also contributing to drought resistance (Kimber 2000; Vietmeyer 1996). Different sorghum are utilized in different ways. The grain can be either sweet or savory and can be used to prepare flat bread, thick porridge, thin porridge, popcorn, vegetable, 'rice' and fodder. The stems, depending on whether they are juicy or not, are used for syrup, beer, liquor, firewood, brooms, baskets, forage and biofuel (Vietmeyer 1996). Sorghum is predominantly selfing, but outcrosses with varying rates ranging from 5 to 40% (Barnaud, et al. 2008; Dje, et al. 2004; Ellstrand and Foster 1983). Furthermore, wild/weedy and cultivated sorghum are infertile with overlapping flowering times.

In the most recent taxonomic treatment (De Wet 1978; Wiersema and Dahlberg 2007) wild sorghum (*S. bicolor* subsp. *verticilliflorum* (Steud.) de Wet ex Wiersema & J. Dahlb*)*, weedy sorghum (*S. bicolor* subsp. *drummondii* (Steud.) de Wet ex Davidse; hybrids between cultivated and wild sorghum) and cultivated sorghum (*S. bicolor* subsp. *bicolor*) are treated as subspecies, within a single species *Sorghum bicolor* (L.) Moench. Within cultivated sorghum (subsp. *bicolor*) five races (guinea, caudatum, bicolor, durra, kafir) and 10 intermediates have been recognized. The races can be identified by differences in their mature spikelets, and can be linked back to the areas and the nomadic people from where they were first cultivated (Harlan and De Wet 1972). Guinea is mostly distributed in west Africa, caudatum in middle and eastern Africa, kafir in southern Africa, durra in Ethiopia, Sudan and India, while bicolor is not associated with any particular area (De Wet and Harlan 1971). Bicolor is known to be the oldest of the races and the most similar to wild sorghum in appearance. Durra, kafir and caudatum have denser panicles, and are cultivated in semi-arid climates, where the rainfall season is short and predictable. Conversely in areas where rainfall can be long and erratic, looser panicles and open glumed guinea types are preferred to avoid grain mold. Durra and caudatum are often used in

breeding programs because of their drought resistance. Dwarfing genes (used to produce short improved sorghums) were originally found among durra varieties (Morris, et al. 2012).

Genetic structure and diversity

The amount of genetic diversity present within a species is often used as a measure for its adaptive ability. Biodiversity is an asset for coping with environmental fluctuations. Factors relevant for the present-day genetic structure and diversity of sorghum are 1) human-mediated effects, 2) mating strategy and 3) gene flow.

Crops are the direct product of human selection on wild plant diversity. Cultivated sorghum is expected, and has been shown, to harbor a lower amount of genetic diversity than wild sorghum (Muraya, et al. 2011b; Mutegi, et al. 2011; Sagnard, et al. 2011). This is normal for cultivated species, which usually undergo a genetic bottleneck during the domestication process, a phenomena that has been observed in sorghum, maize, wheat, rice and soybean (Eyre-Walker, et al. 1998; Guo, et al. 2010; Haudry, et al. 2007; Zhu, et al. 2007). Human selection for desirable traits is a continuous process, shaping the diversity and structure of cultivated species. In the words of Darwin:

> "Whatever part or character is most valued- whether the leaves, stems, tubers, flowers, fruit, or seed of the plant…-that character will almost invariably be found to present the greatest amount of difference both in kind and degree. And this result may be safely attributed to man having preserved variations which were useful to him, and neglecting the others" (Darwin 1868: p.220).

Compared to outcrossing species, self-fertilizing species are more prone to strong population differentiation (due to genetic drift) and loss of genetic diversity (due to reduced heterozygosity levels). The long-term evolutionary effects of a self-fertilizing mating strategy, however, are poorly documented (Takebayashi and Morrell 2001), because molecular marker based estimates of genetic diversity may not be representative of the quantitative genetic variation (Cheverud, et al. 2002; Storfer 1996).

Gene flow between cultivated and wild (or weedy) plants has occurred for centuries. Most cultivated plants mate with their wild relatives within some portion of their geographical range (Ellstrand, et al. 1999). Many cases of gene flow have been recorded where the habitats of wild (or weedy) and cultivated sorghum overlap (Adugna, et al. 2012; Arriola and Ellstrand 1996, 1997; Barnaud, et al. 2009; Dje, et al. 2004; Muraya, et al. 2011a; Mutegi, et al. 2012; Mutegi, et

al. 2010). Gene flow between wild (or weedy) and cultivated sorghum can have both positive and negative (from a human point of view) consequences. It is acknowledged that wild-crop gene flow is a source of introducing genetic diversity into crop populations (Jarvis and Hodgkin 2002). On the other hand, gene flow has also been implied as a force for the creation of aggressive weeds (Ellstrand and Schierenbeck 2000). One such example is Johnson grass (*Sorghum halepense* (L) Pers.), listed as one of the world's 10 worst weeds (Holm 1969), which is a cross between cultivated (*S. bicolor* subsp. *bicolor*) and a wild sorghum relative (*S. propinquum* (Kunth) Hitchc.) (Monaghan 1979). Another fear is that genes from improved or genetically modified crops may enter into wild populations and infer fitness advantages on already troublesome weeds (Conner, et al. 2003; Ellstrand and Hoffman 1990; Gepts and Papa 2003; Stewart, et al. 2003). Gene flow is usually higher in the crop-wild direction, due to the size difference between crop and wild populations, and higher outcrossing rates of wild sorghum compared to cultivated sorghum (Muraya, et al. 2011a). This has raised conservation concerns regarding the genetic swamping, and in the worst case, extinction of wild populations (Ellstrand, et al. 1999; Gepts and Papa 2003).

In this study the genetic structure and diversity of sorghum landraces will be investigated at a continental scale (involving cultivated sorghum from throughout Africa), a country scale (involving cultivated sorghum from throughout Tanzania) and a local scale (involving cultivated and wild/weedy sorghum from a single village). We sought to address the following questions:

a) *To what degree is the genetic diversity of sorghum structured based on*
    1) *geography*
    2) *climate/race*
    3) *landrace/grain color*
b) *In what ways do human cultivation practices and mating system influence the genetic diversity and structure of cultivated sorghum?*
c) *Does gene flow occur between landraces and between landraces and wild/weedy sorghum?*

MATERIALS AND METHODS

Plant material

The sorghum material included in this study (Table 1 Apx1) represents a continental scale (Africa), a country scale (Tanzania) and a local scale (Hombolo). For the continental scale study we obtained 41 accessions of cultivated sorghum from 13 countries in Africa and one location in India (Figure 1) from ICRISAT (International Crops Research Institute for the Semi-Arid Tropics, Hyderabad, India). Racial information was obtained for 30 of the ICRISAT accessions (from the following website; https://www.soils.org/publications/cs/abstracts/49/5/1769). For the country scale study we acquired 42 accessions of cultivated sorghum from 12 provinces in Tanzania (Figure 2) from the NPGRC (National Plant Genetic Resource Institute, Arusha, Tanzania). For the local scale study 86 accessions, including both cultivated and wild/weedy sorghum, were collected in Hombolo, a village situated close to Dodoma, the capital of Tanzania. Sorghum is a common crop in this area because of the semi-arid climate. Five households were selected with the assistance of the Hombolo Agricultural Research Institute. The households are named 1-5 according to their geographical placement in Hombolo (Figure 2). At each household we collected panicles and leaf samples (placed directly in silica gel) from five plants of the most commonly grown sorghum landraces, and one plant from the less common landraces. Panicles and leaf samples were also sampled from five wild/weedy sorghum plants found either interspersed, or adjacent to the crop fields at each household. The wild/weedy sorghum was identified by its small grains and loose panicles (Figure 3). Additionally, five panicles and leaf samples were collected from a population of wild/weedy sorghum between household 4 and household 5. Sampling locations were ascertained using GPS coordinates. Each farmer was interviewed and asked questions regarding their sorghum crop. For the accessions collected in Hombolo we used the landrace names provided by the farmers (Figure 3). During sampling we made no attempt to distinguish between wild (subsp. *verticilliflorum)* or crop-wild hybrids (subsp. *drummondii)*, and refer to the whole non cultivated pool as 'wild', as done in the study by Mutegi, et al. (2011). A copy of each landrace sample was deposited at the NPGRC for conservation and future utilization, and the grain colors of the landraces were registered. The plant material was brought out of Tanzania under a phytosanitary certificate and standard material transfer agreement (ftp://ftp.fao.org/ag/agp/planttreaty/agreements/smta/SMTAe.pdf).

Climate data was retrieved from the GENESYS database (www.genesys-pgr.org) for all the geo-referenced accessions.

DNA extraction

Sorghum grains were germinated in a greenhouse (25°C and 12 hrs. daylight) at the University of Oslo. For each accession 10-20 cm long leaves were cut and dried in silica gel for 2-3 weeks. Dried leave material from each accession were placed in an eppendorf tube together with carbide beads and crushed in a mixer mill MM301 (Retsch, GmbH & Co., Haan) for 2-3 min at 20 Hz, in preparation for DNA extraction. DNA was extracted from 169 sorghum accessions using the E.Z.N.A SP Plant Mini kit (Omega Bio-Tek, Norcross) without any modifications of the manufacture's manual. A Nanodrop 1000 spectrophotometer (Thermo Scientific, Wilmington) was used to assess the quality and concentration of the extracted DNA. All DNA extractions with concentrations above 20 ng/µL were diluted 5x for the microsatellite analysis.

Microsatellite analysis

Twenty microsatellites (Table 1) were selected from previously published primers (Bhattramakki, et al. 2000; Brown, et al. 1996; Kong, et al. 2000; Mutegi, et al. 2011; Schloss, et al. 2002; Taramino, et al. 1997), out of which 18 were picked from a published sorghum microsatellite kit (Billot, et al. 2012). In order to pool and distinguish microsatellites during electrophoresis the M13 tailing approach was used (Schuelke 2000). In this method the forward primer of each microsatellite is tagged with a fluorescently labeled M13 primer (TGTAAAACGACGGCCAGT).

Figure 1 **Map of Africa and India showing the collection sites (green dots) of sorghum accessions obtained from ICRISAT (Hyderabad, India). Countries from which sorghum material was obtained are colored in yellow.**

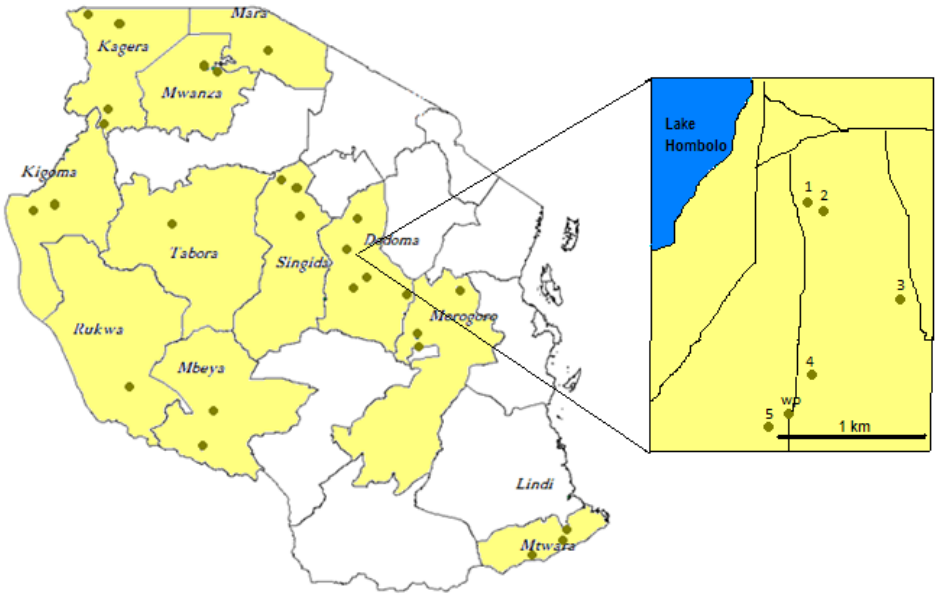


Figure 2 **Map of Tanzania showing the collection sites (green dots) of sorghum accessions obtained from NPGRC (Arusha, Tanzania), and collection sites of sorghum accessions sampled from five households (numbered 1-5) and one wild/weedy population (wp) in Hombolo, Dodoma (insert). Provinces from which sorghum material was obtained are colored in yellow.**

Four fluorescently-labeled M13 primers were used; FAM (blue), NED (yellow), VIC (green), and PET (red). All unlabeled primers, as well as FAM- labeled M13 primers, were purchased from IDT (Integrated DNA Technologies, Leuven). NED-, VIC- and PET- labeled M13 primers were purchased from Applied Biosystems (Foster City). A total volume of 10µL PCR reactions were made containing 1 µL 10x CoralLoad PCR buffer (Qiagen, Hilden), 1 µL 2 mM dNTP, 0.2 µL 5 µM forward primer, 0.8 µL 5 µM reverse primer, 0.8 µL 5 µM fluorescent-labeled M13 primer (Table 1), 0.05 µL HotStar Taq *Plus* DNA polymerase (Qiagen, Hilden), 4.15 µL mqH_2O, 2 µL diluted DNA template. The PCR reactions were amplified in a DNA Engine Tetrad 2 Peltier Thermal Cycler (Bio-Rad Laboratories, Hertfordshire) using the following conditions: An initial denaturation step at 95°C for 5 min followed by 30 cycles, each consisting of a denaturation step at 95°C for 30 sec, an annealing step at x°C for 45 sec (temperatures specified for each marker in Table 1), and an extension step at 72°C for 45 sec. This was followed by an additional eight cycles consisting of a denaturation step at 95°C for 30 sec, an annealing step at 53°C for 45 sec, and an extension step at 72°C for 45 sec. The program ended with a final annealing step at 72° C for 30 min. The success of the PCR reactions was checked using electrophoresis with 1% agarose gel and GelRed (Biotium, Hayward) as the DNA stain. 5 µL of GelRed was used per 80 µL of agarose gel, and 2-3 µL of PCR products were loaded onto the gel. The PCR products were diluted 10x, and five microsatellites with different dyes and non-overlapping lengths were pooled (4µL of the two microsatellites tagged with FAM and 3µL of each of the microsatellites tagged with VIC, NED, and PET). Further, 1µL of this mixture was mixed with 8.85 µL Hi-Di™ formamide and 0.15 µL GeneScan™ 500 (-250) LIZ size standard (both from Applied Biosystems) in a total volume of 10µL, which was denatured at 95°C for 5 min, and then applied to an ABI 3730 DNA analyzer (Applied Biosystems), in order to measure the length of the microsatellite fragments. For each 96 welled plate we included two negative controls, five internal replicates and two replicates between different runs.

Table 1 **Summary of microsatellite primers used to analyze the sorghum material. Developed by- indicates where the microsatellites were first published. K- indicates which microsatellites were chosen (marked with X) from the microsatellite kit (Billot, et al. 2012). F refers to the forward primer and R refers to the reverse primer. Rep.m.- is the repeat motif. Length- is the bp range within which alleles were found in this study. An. T.- is the annealing temperature used during PCR. M13. L.- refers to the dye used for each microsatellite. Microsatellites that were excluded from the final analysis are marked with *.**

Locus	K	Developed by	Primer sequence 5`-3`	Rep.m.	Length	An.T.	M13.L.
sb5-206 (xgap206)	**X**	**Brown, et al. (1996)**	**F:ATTCATCATCCTCATCCTCGTAGAA R:AAAAACCAACCCGACCCACTC**	**(AC)13(AG)20**	**97-175**	**55**	**FAM**
sb5-236	**-**	**Brown, et al. (1996)**	**F:GCCAAGAGAAACACAAACAA R:AGCAATGTATTTAGGCAACACA**	**(AG)20**	**180-220**	**55**	**NED**
Xcup02	**X**	**Schloss, et al. (2002)**	**F:GACGCAGCTTTGCTCCTATC R:GTCCAACCAACCCACGTATC**	**(GCA)6**	**207-222**	**55**	**VIC**
Xcup14	**X**	**Schloss, et al. (2002)**	**F:TACATCACAGCAGGGACAGG R:CTGGAAAGCCGAGCAGTATG**	**(AG)10**	**213-266**	**58**	**PET**
Xcup61	**X**	**Schloss, et al. (2002)**	**F:TTAGCATGTCCACCACAACC R:AAAGCAACTCGTCTGATCCC**	**(GAG)7**	**210-220**	**58**	**VIC**
Xtxp15	**X**	**Kong, et al. (2000)**	**F:CACAAACACTAGTGCCTTATC R:CATAGACACCTAGGCCATC**	**(TC)16**	**218-242**	**55**	**PET**
Xtxp40	**X**	**Kong, et al. (2000)**	**F:CAGCAACTTGCACTTGTC R:GGGAGCAATTTGGCACTAG**	**(GGA)7**	**140-158**	**55**	**FAM**
Xtxp57	**X**	**Bhattramakki, et al. (2000)**	**F:GGAACTTTTGACGGGTAGTGC R:CGATCGTGATGTCCCAATC**	**(GT)21**	**253-258**	**55**	**PET**
Xtxp289	**-**	**Bhattramakki, et al. (2000)**	**F:AAGTGGGGTGAAGAGATA R:CTGCCTTTCCGACTC**	**(CCT)16(AGG)6**	**275-351**	**58**	**FAM**
Xtxp295	**X**	**Bhattramakki, et al. (2000)**	**F:AAATCATGCATCCATGTTCGTCTTC R:CTCCCGCTACAAGAGTACATTCATAGCTTA**	**(TC)19**	**160-210**	**55**	**NED**
Xtxp12	**X**	**Kong, et al. (2000)**	**F:AGATCTGGCGGCAACG R:AGTCACCCATCGATCATC**	**(CT)22**	**181-225**	**58**	**VIC**
gpsb123	**X**	**Mutegi, et al. (2011)**	**F:ATAGATGTTGACGAAGCA R:GTGGTATGGGACTGGA**	**(AC)7(GA)5**	**299-314**	**55**	**FAM**
sbAGB02	**X**	**Taramino, et al. (1997)**	**F:CTCTGATATGTCGTTGTGCT R:ATAGAGAGGATAGCTTATAGCTCA**	**(AG)35**	**90-160**	**55**	**FAM**
mSbCIR283	**X**	**Billot, et al. (2012)**	**F:TCCCTTCTGAGCTTGTAAAT R:CAAGTCACTACCAAATGCAC**	**(CT)8 (GT)8**	**130-180**	**55**	**FAM**
Xtxp320	**X**	**Bhattramakki, et al. (2000)**	**F:TAAACTAGACCATATACTGCCATGATAA R:GTGCAAATAAGGGCTAGAGTGTT**	**(AGG)20**	**265-307**	**58**	**FAM**
Xtxp141	**X**	**Bhattramakki, et al. (2000)**	**F:TGTATGGCCTAGCTTATCT R:CAACAAGCCAACCTAAA**	**(GA)23**	**149-185**	**55**	**NED**
Xtxp278	**X**	**Bhattramakki, et al. (2000)**	**F:GGGTTTCAACTCTAGCCTACCGAACTTCCT R:ATGCCTCATCATGGTTCGTTTTGCTT**	**(TTG)12**	**255-302**	**53**	**FAM**
sb4-72* (xgap72)	X	Brown, et al. (1996)	F:TGCCACCACTCTGGAAAAGGCTA R:CTGAGGACTGCCCCAAATGTAGG	**(AG)16**	177-215	55	NED
mSbCIR262*	X	Mutegi, et al. (2011)	F:GCACCAAAATCAGCGTCT R:CCATTTACCCGTGGATTAGT	**(CATG)3.25**	220-245	55	VIC
Xtxp136*	X	Kong, et al. (2000)	F:GCGAATAGCATCTTACAACA R:ACTGATCATTGGCAGGAC	**(GCA)5**	255-265	55	PET

The microsatellite fragment sizes were viewed in GeneMapper version 3.7 (Applied Biosystems) based on migration relative to the internal sizing standard. Alleles were chosen using both automated allele scoring (implemented in GeneMapper) and manual editing. Three markers were excluded from further analysis; mSbCIR262 and sb4-72 (x-gap72) were removed due to a high amount of missing data, 50% and 20% respectively and Xtxp136 because the alleles could not be unambiguously scored.

Data analysis

The final allelic data was copied from GeneMapper and formatted into 1) a presence/-absence data matrix and 2) a data matrix with allele sizes in base pairs (bp). Accessions with missing data for five or more (> 30%) markers were removed from the dataset (eight accessions, see Table Apx1) together with all replicates (after confirming that they represented identical patterns). The data matrix with allele sizes in bp was used as input file for the software CONVERT (Glaubitz 2004), which was used to create input files for STRUCTURE version 2.2, (Pritchard, et al. 2000), the web version of GENEPOP (http://genepop.curtin.edu.au/), ARLEQUIN (Excoffier, et al. 2005) and HP-Rare version 1.2 (Kalinowski 2005).
The full dataset (161 sorghum accessions) was subdivided into smaller datasets for further analysis. Dataset 'Africa' consisted of 41 cultivated accessions from throughout Africa. Dataset 'Tanzania' consisted of 42 cultivated accessions from throughout Tanzania. Dataset 'Hombolo' consisted of 52 cultivated accessions collected from five households in Hombolo, Tanzania. Dataset 'Wild' consisted of 26 'wild' sorghum accessions collected from five households and one 'wild' population in Hombolo, Tanzania. To analyze the cultivated accessions from Hombolo together with 'wild' accessions, dataset 'Hombolo' was analyzed together with dataset 'Wild' (altogether 78 accessions). To analyze the accessions from Hombolo in a larger scale context, dataset 'Hombolo' was analyzed together with dataset 'Tanzania' (altogether 94 accessions). For details on the definition of groups and the specific data analyses performed on each dataset see Table Apx2.

Genetic diversity was assessed using expected heterozygosity (He), observed heterozygosity (Ho), allelic richness (Rs), private allelic richness (PRs) and inbreeding coefficient (Fis) as diversity measures. Fis calculates the proportion of the population that is autozygous (i.e. the alleles that are identical by descent and originated from the same parent) (Futuyma 2009). Fis

ranges from 0 (the population is mating randomly) to 1 (the population is completely autozygous). He, Ho and Fis were computed using ARLEQUIN version 3.1. Rs and PRs were calculated using the software HP-Rare version 1.2, which uses a rarefaction method that adjusts for uneven sample size. Furthermore, linkage disequilibrium was calculated, with the null hypothesis "genotypes at one locus are independent from genotypes at the other locus", using the web version of the software GENEPOP.

Genetic divergence and gene flow between populations were assessed by Analysis of Molecular Variance (AMOVA), fixation index based on allele sizes (Rst) and number of migrating accessions (Nm). AMOVA and Rst were calculated using ARLEQUIN version 3.1. AMOVA tests the significance of covariance components using non-parametric permutation procedures, and explains how much variation is partitioned among populations and within populations. The AMOVA analysis yields an overall Fst value which was used to summarize the AMOVA output. Fst and Rst both measure the genetic difference between two populations (Rst also takes into account allele lengths) and range from 0 (the two populations have the same allele frequencies) to 1 (the two populations are fixed for different alleles). As pairwise Rst and pair wise Fst yielded similar results, only the pairwise Rst values are shown. The number of migrating accessions (Nm) was calculated in GENEPOP using a method developed by Slatkin (1985) based on the amount of private alleles found in each population. He discovered a linear relationship between the amount of private alleles and the number of migrating accessions, as migration increases the number of private alleles decrease and vice versa (Slatkin 1985). Comparisons of mean Rst, He, PRs and Rs values were tested for significance among different groups using a Wilcoxon Signed Rank test (Wilcoxon 1945), computed in PAST version 2.13 (Hammer, et al. 2001). Calculations of Nm, Rst and linkage disequilibrium were only done for the accessions from Hombolo.

In order to view genetic structure, three approaches were used; ordination, Bayesian clustering and neighbor joining analysis. For the ordination approach two methods were used. A Principal Coordinate (PCO) analysis was constructed using the presence/absence data matrix and the software PAST version 2.13 with Dice's coefficient as similarity measure. The eigenvalues from PAST were used to create a plot in R version 2.11 (R Development Core Team 2010). A Principal Component Analysis (PCA) was constructed in R version 2.11 using the data matrix

with allele sizes in bp and the R package ADEGENET (Jombart 2008). Ordination shows the relationship between accessions without any assumptions or hierarchy. PCO uses discrete measurements, such as the presence or absence of an allele and a distance or similarity measure, such that the distance between all the points in the graph corresponds to the distances between accessions in the dataset. PCA, on the other hand, uses continuous measurements (allele sizes in our case) in order to view structure of the data (Pielou 1984). For Bayesian clustering the software STRUCTURE version 2.2 was used through the Bioportal computer service (University of Oslo; http://www.bioportal.uio.no). STRUCTURE is a model based program that assigns accessions probabilistically to one of K clusters in such a way as to achieve Hardy Weinberg equilibrium. Groups (K) between 1 and 9 were tested in order to find the optimal number of groups for the dataset. The program was run using the following settings: 10^6 iterations, followed by a burnin of 10^5, using the admixture model, which allows for shared ancestry between different clusters and the correlated allele frequency model, which allows for similar allele frequencies in different groups. For each K, 10 rounds were run. The optimal K was chosen by summarizing the STRUCTURE outputs using a collection of R functions implemented in STRUCTURE-SUM (Ehrich 2006). In addition to providing a summary of the posterior (logarithmic) probabilities (ln P(D)), STRUCTURE-SUM plots the similarity coefficient (Nordborg, et al. 2005) and delta K of the data (Evanno, et al. 2005). Plots of the STRUCTURE groups were made with the software *distruct* version 1.1(Rosenberg 2003) using the output files from STRUCTURE. As sorghum is predominantly selfing populations probably deviate significantly from Hardy Weinberg equilibrium. With these considerations, a program similar to STRUCTURE, but developed for self-fertilizing species InStruct (Gao, et al. 2007) was applied for some of the datasets. The results from InStruct were similar to those from STRUCTURE and only the results from STRUCTURE are presented. Neighbor joining (NJ) analysis (Saitou and Nei 1987) was calculated in PAST version 2.13 using the presence/absence data matrix, Dice's coefficient as similarity measure and the default rooting (on the last branch added during tree construction). The NJ tree was edited in FigTree version 1.3.1 (Rambaut 2008). To explore possible associations between geographical distance and genetic structure, Mantel tests were calculated in PAST version 2.13. A Mantel test is calculated using two dissimilarity matrices. The null hypothesis is that the distances in matrix A are independent of the corresponding distances in matrix B (Bonnet and Van de Peer 2002; Dietz 1983; Mantel 1967). Dice's

coefficient was used for estimating genetic similarity between allele sizes in matrix A and Euclidean distance was used for estimating the distance between geographical coordinates in matrix B.

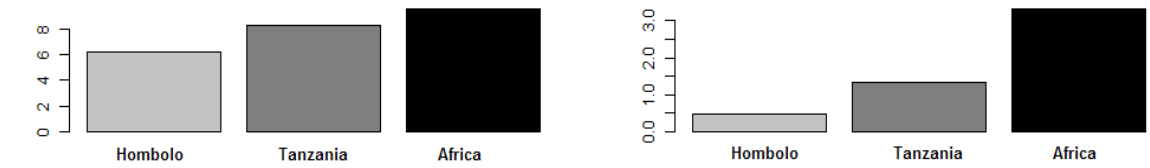
Figure 3 **Panicles (or parts of panicles) of eight cultivated sorghum landraces and two 'wild' morphs collected in Hombolo, Tanzania, 3-5 June 2011.**

RESULTS

All the microsatellite loci were polymorphic across the 161 sorghum accessions in this study. The number of alleles per locus ranged from four alleles (Xcup61, Xcup02) to 29 alleles (sb5-206) with an average of 13.11 alleles per locus. The Xcup series had a lower number of alleles, likely because they are located inside or closer to genes (Schloss, et al. 2002). The percentage of missing data per locus (for the loci included in the final data analysis) ranged from 1.6% (mSbCIR283) to 8.9% (Xtxp15), with an average of 3.15% missing data per locus. The observed heterozygosity (Ho) ranged from 1.8% (gpsb123) to 20% (Xtxp295), with an average of 10.3% per locus.

Africa, Tanzania and Hombolo

Results from the complete dataset showed that there was a significant ($p \leq 0.05$) increase in allelic richness (Rs) and private allelic richness (PRs) for cultivated sorghum, with increasing geographical scale (Figure 4).



A) B)

Figure 4 A) Barplot of mean allelic richness and B) mean private allelic richness adjusted for sample size, for sorghum accessions based on 17 microsatellite loci. Hombolo- refers to 52 cultivated sorghum accessions collected from five households in Hombolo, Tanzania. Tanzania- refers to 42 cultivated sorghum accessions from throughout Tanzania and Africa- refers to 41 cultivated sorghum accessions from throughout Africa. Differences between all means were significant ($p<0.05$) using a Wilcoxon Signed Rank test.

The continental (Africa) scale dataset (41 cultivated accessions) had an average of 1.22% Ho per locus. The country (Tanzania) scale dataset (42 cultivated accessions) had an average of 1.25% Ho per locus, while the local (Hombolo) scale dataset (52 cultivated accessions) had an average of 14% Ho per locus.

Africa

For 41 cultivated accessions from throughout Africa, AMOVA analysis showed that the greatest genetic differentiation was found between the groups identified by STRUCTURE (Fst=0.16), then by country (Fst=0.11), temperature (Fst=0.08), race (Fst=0.07), precipitation (Fst=0.04) and finally, by grain color (Fst=-0.01) (Table 2, Table Apx3). Four STRUCTURE groups (Figure 5A) were chosen based on the output from STRUCTURE-SUM (Figure 5B). As the optimal K was not evident by the ln P(D) plot, the delta K plot was used to choose K. The four groups mostly displayed a geographical structure, a result supported by a Mantel test (R=0.36, $p \leq 0.05$).The STRUCTURE group 'West' mainly included accessions from Mali and Nigeria. 'Northeast' mainly included accessions from Somalia, Ethiopia and Sudan, whilst 'East' mainly included accessions from Tanzania, Kenya and Uganda. 'South' mainly included accessions from South Africa, Lesotho, Botswana, Malawi and Zimbabwe. Exceptions to the geographical structure were the accessions from Sudan, which were found in every group. A South African accession was found in 'Northeast', and 'West', a Ugandan, and Tanzanian accession in 'West' and a Nigerian, Indian and Zimbabwean accession in 'Northeast'. A racial structure was also seen within the STRUCTURE groups, although this did not account for as much of the genetic variation as did the geographical structure (Table 2, Table Apx3). 'Kafir' was found exclusively in the southern group, 'guinea' was mostly found in the western group, 'durra' was mostly found in the northeastern group, 'caudatum' was mostly found in the eastern and northeastern groups and 'bicolor' was found in the western and northeastern groups (Figure 5A).

Table 2 **Overall Fst, Fis and p values from AMOVA for cultivated and 'wild' sorghum, based on 17 microsatellite markers. Africa- refers to 41 accessions from throughout Africa, Tanzania- refers to 42 accessions from throughout Tanzania and Hombolo- refers to 26 'wild' and 52 cultivated sorghum accessions, collected from five households in Hombolo, Tanzania. For the African accessions populations were defined based on 1) STRUCTURE analysis (K=4), 2) geography (11 countries), 3) race (guinea, bicolor, caudatum, durra, kafir and intermediates), 4) grain color (red, white), 5) mean temperature at the collection sites (temp) and 6) mean annual precipitation from the collection sites (prec). For the Tanzanian accessions populations were defined based on 1) groups identified by STRUCTURE (K=5), 2) geography (11 provinces), 3) grain color (red, white), 4) mean temperature at the collection sites (temp) and 5) mean annual precipitation from the collection sites (prec). For the Hombolo accessions populations were defined based on 1) STRUCTURE analysis (K=5), 2) household affiliation (households 1-5), 3) eight landrace (landraces), 4) grain color (red,white) and 5) cultivated versus 'wild' sorghum (cultivated,'wild'). Significance tests consisted of 1640 permutations.**

Datasets	Overall Fst	Overall Fis	p
Africa (STRUCTURE)	0.16	0.97	0.001
Africa (geography)	0.11	0.97	0.001
Africa (grain color)	-0.01	0.98	0.450
Africa (race)	0.07	0.98	0.990
Africa (temp)	0.08	0.97	0.100
Africa (prec)	0.04	0.97	0.100
Tanzania (STRUCTURE)	0.24	0.78	0.001
Tanzania (geography)	0.19	0.80	0.001
Tanzania (grain color)	0.12	0.82	0.001
Tanzania (temp)	0.06	0.83	0.900
Tanzania (prec)	0.04	0.83	0.900
Hombolo (STRUCTURE, no 'wild')	0.33	0.57	0.001
Hombolo (households, no 'wild')	0.15	0.68	0.001
Hombolo (landraces)	0.14	0.69	0.002
Hombolo (grain color)	0.19	0.70	0.001
Hombolo (cultivated, 'wild')	0.03	0.72	0.004

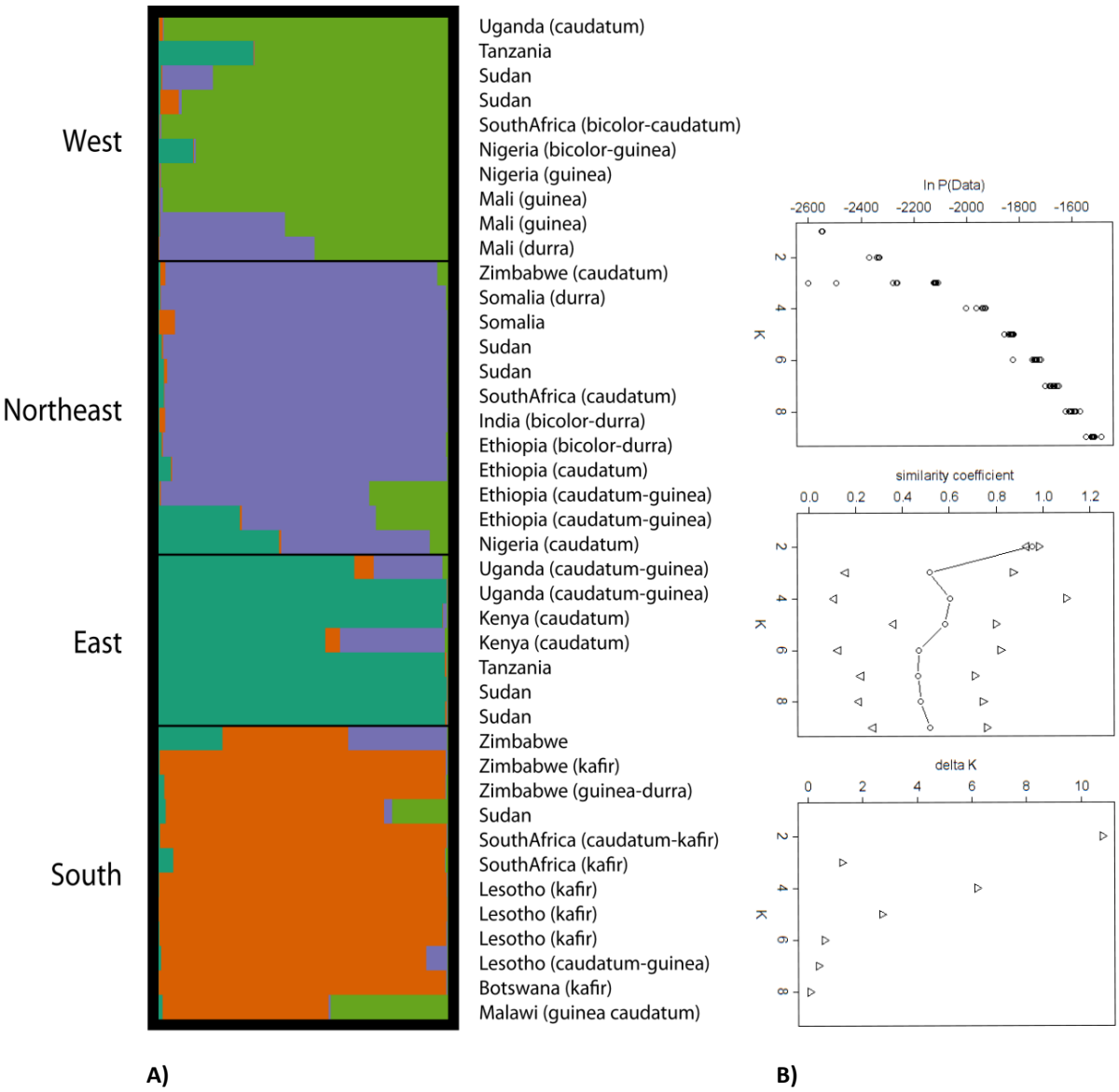


Figure 5 **A) Plot of STRUCTURE results (K=4) for 41 accessions of cultivated sorghum from throughout Africa, based on 17 microsatellites. The race (kafir, guinea, caudatum, durra, bicolor and intermediates) of the accessions (when available) is shown in brackets. The groups (K) are represented by different colors. The segmentation of the horizontal pillars shows with what percentage an accession is placed within which groups B) Plot of the output from STRUCTURE-SUM including a summary, for k=1-9, of the logarithmic probability (ln P(D)), the similarity coefficient and delta K.**

Tanzania

For 42 cultivated accessions from throughout Tanzania, AMOVA analysis showed that the greatest genetic differentiation was found between the groups identified by STRUCTURE (Fst=0.24), then by grain color (Fst=0.12) and lastly, by temperature (Fst=0.06) and precipitation (Fst=0.04) (Table 2, Table Apx3). Five STRUCTURE groups (Figure 6A) were chosen based on the output from STRUCTURE-SUM. As the optimal K was not evident by the ln P(D) plot, the delta K plot was used to choose K (Figure 6B). The five groups identified by STRUCTURE displayed a geographical distribution (Figure 7) supported by PCA (Figure Apx1) and a Mantel test ($R=0.34$, $p \leq 0.05$). The STRUCTURE group 'North' included accessions from northern provinces (Mara and Kagera). 'Northwest' included accessions from a northern province (Mwanza) and a western province (Kigoma). 'Central Northwest' and 'Central Southwest' included mainly accessions from several central (Dodoma, Singida and Morogoro) and western (Rukwa, Mbeya and Kigoma) provinces. 'Southeast' included mainly accessions from a southern province (Mtwara) and from the coast (exact locations for these accessions are not known). Exceptions to the geographical structure were three accessions from Kigoma which were distributed across three different groups ('Northwest', 'Central Northwest' and 'Southeast'), and some accessions from central and western provinces (Tabora, Kigoma, Rukwa, Mbeya, Morogoro, Singida and Dodoma) which were included in 'Southeast' (Figure 6A).

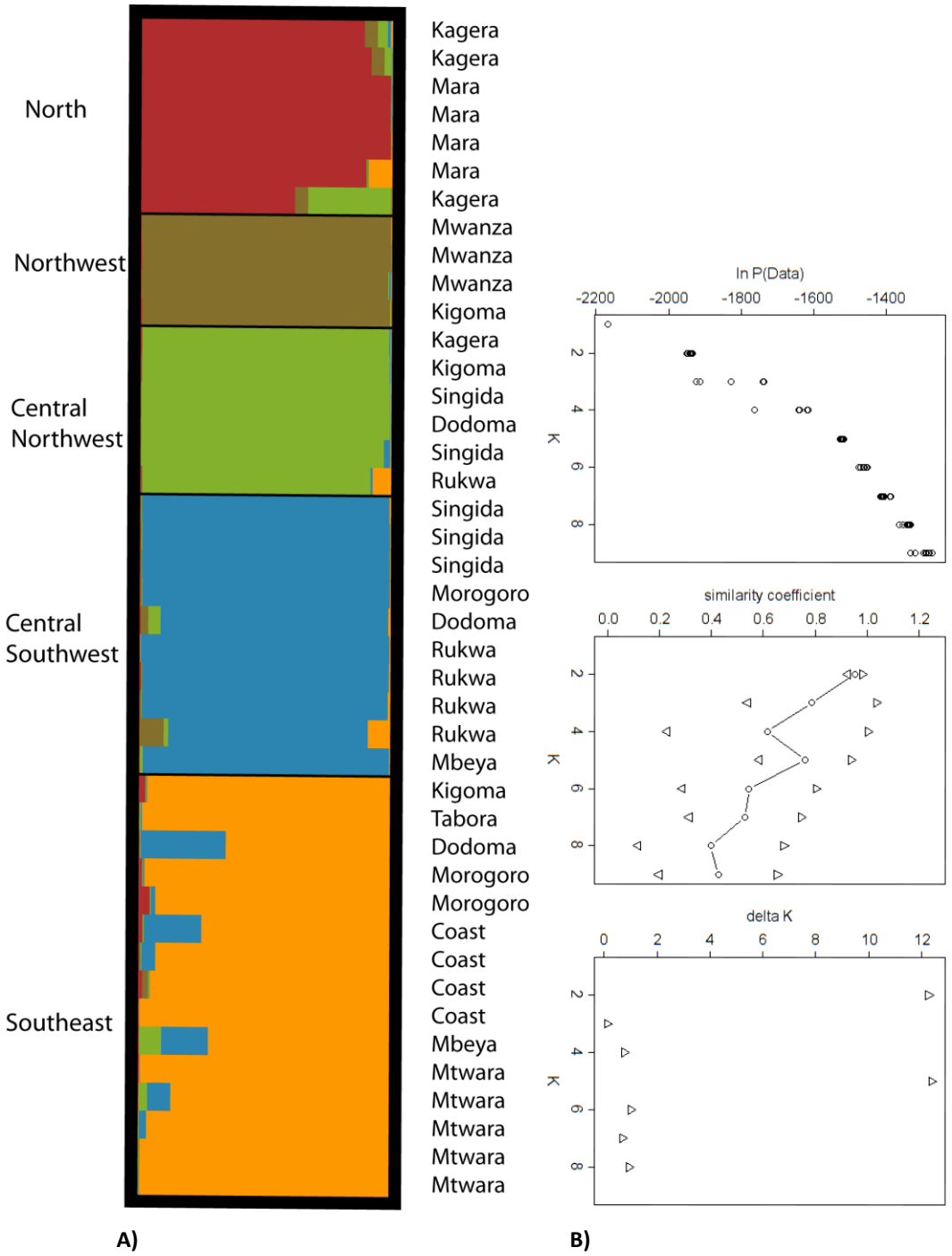


Figure 6A) Plot of the STRUCTURE results (K=5) for 42 cultivated sorghum accessions from throughout Tanzania, based on 17 microsatellites. The groups (K) are represented by different colors. The segmentation of the horizontal pillars shows with what percentage an accession is placed within which groups. B) Plot of the output from STRUCTURE-SUM including a summary, for K=1-9, of the logarithmic probability (ln P(D)), the similarity coefficient and delta K.

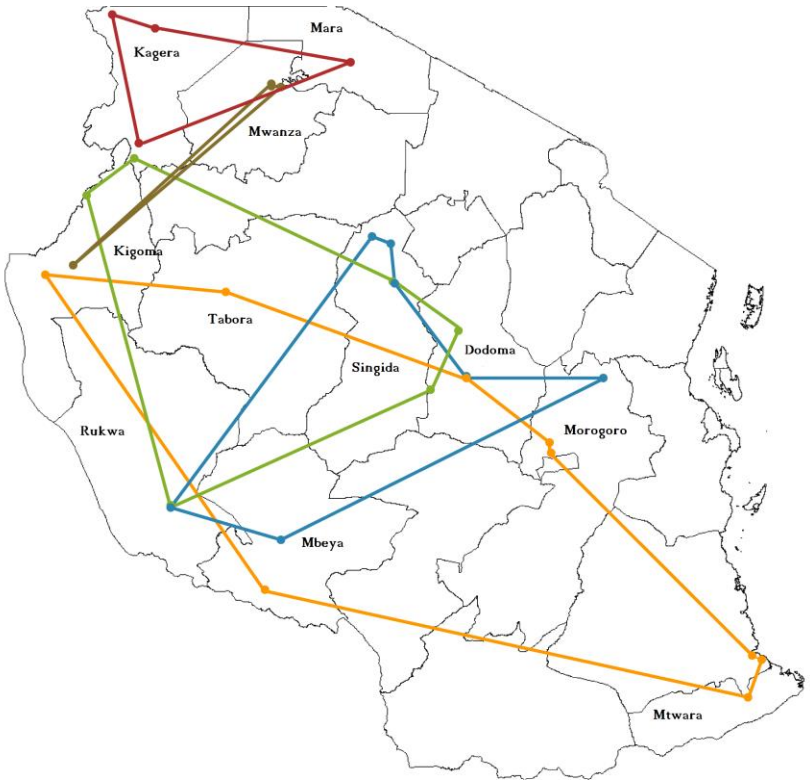


Figure 7 Map of Tanzania with a graphical presentation of the five groups identified by STRUCTURE, for 42 sorghum accessions from throughout Tanzania, based on 17 microsatellite markers. Accessions included in the STRUCTURE groups are connected by colored polygons which correspond to the colors used for the STRUCTURE groups in Figure 6A.

Hombolo

Eight different landraces (Figure 3) were sampled from five households in Hombolo; 2-5 landraces per household. The landraces 'white lugugu' and 'black lugugu' were the main crop in household 1,2,4 and 5, whereas 'namata' was the main crop in household 3. The landraces 'roma' and 'magaje' were cultivated (and intermingled with the main crop) in household 3,4 and 5, while 'wawa', 'limondigua' and 'sandala' were cultivated in just one household each (Table Apx4). Lugugu is a name for sorghum in the native language (gogo), and white lugugu and black lugugu are thought to have been cultivated in the area for a long time. Roma on the other hand, was given its name because it arrived with Italian missionaries. During interviews (Table Apx 5) it was discovered that all the five farmers considered the 'wild' sorghum to be a problem due to its presence amongst the cultivated sorghum, and because it was difficult to identify them as

'wild' until maturity. Also discovered during interviews was that farmers obtained seeds from their neighbors, or the recycled grains from their own crops, and that grains of different landraces were commonly mixed together before sowing. When asked their opinion on improved varieties of sorghum, all the farmers preferred local landraces because of storage and palatability.

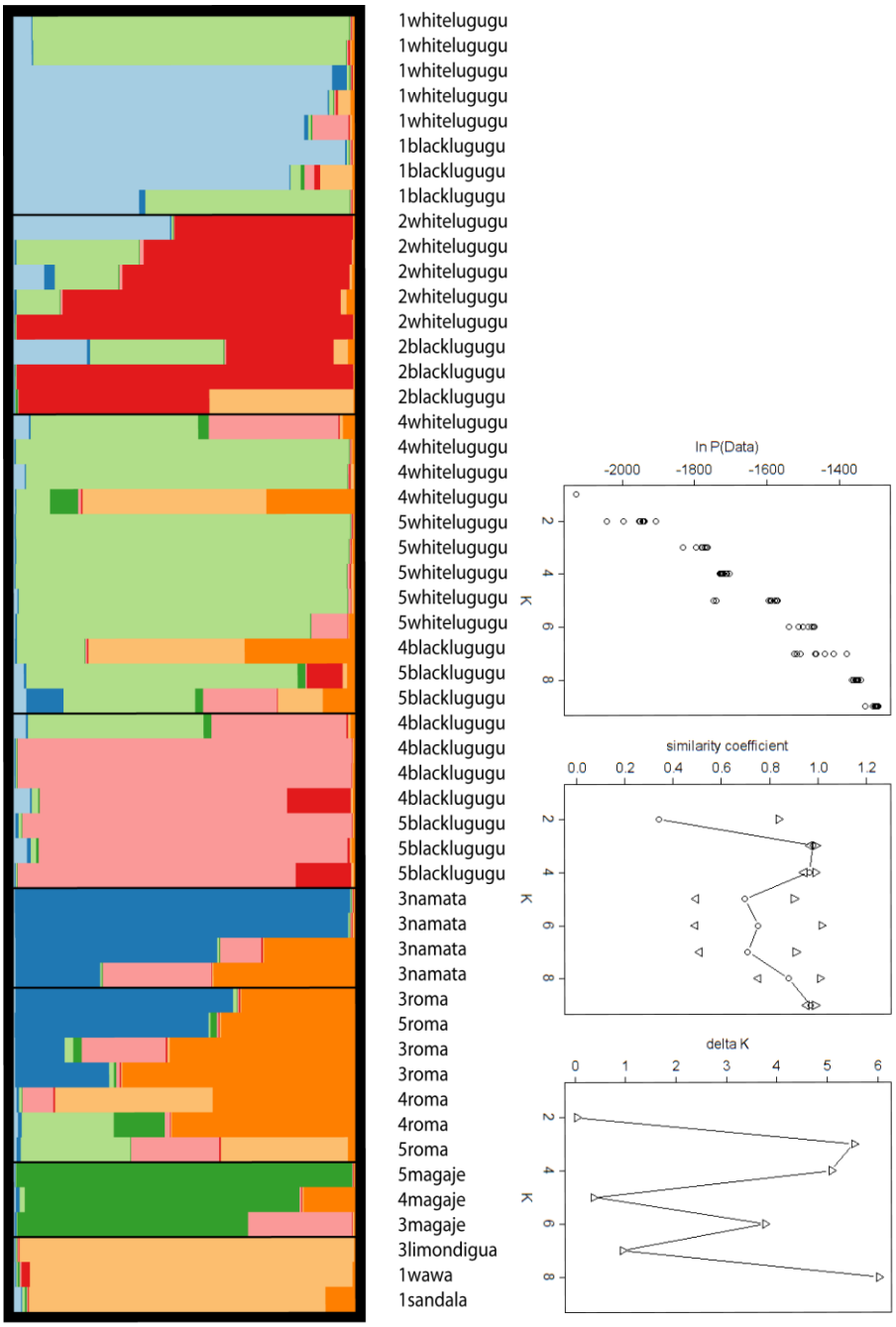
Landraces

For 52 cultivated accessions from Hombolo, AMOVA analysis showed that the greatest genetic differentiation was found between groups identified by STRUCTURE (Fst=0.33), followed by grain color (Fst=0.19), then by households (Fst=0.15) and finally, by different landraces (Fst=0.14; Table 2, Table Apx3). A Mantel test revealed a significant correlation between genetic structure and geographical distance ($R=0.22$ $P \leq 0.05$) for landraces, however, there was *no* significant correlation between geographical distance and genetic structure in 'wild' sorghum ($R=0.07$, $p=0.053$). Eight STRUCTURE groups (Figure 8A) were chosen based on the output from STRUCTURE-SUM. As the optimal K was not evident by the ln P(D) plot, the delta K plot was used to choose K (Figure 8B).The groups mainly displayed differentiation between landraces (also seen in the PCA; Figure Apx2), but also showed some degree of household differentiation. The first four groups consisted of white lugugu and black lugugu (Figure 8A light blue, red, pale green and pink group). The remaining four groups were largely comprised of magaje (green), roma (mustard and blue), namata (blue) and a joint group including wawa, limondigua and sandala (peach). The light blue group was comprised of all the white lugugu and black lugugu accessions from household 1, with the exception of two white lugugu accessions which joined the pale green group. The red group was comprised of white lugugu and black lugugu from household 2. Most of the white lugugu from household 4 and 5 (pale green) and most of the black lugugu from household 4 and 5 (pink) formed their own groups, except for some black lugugu accessions which joined the pale green group. The intermixed state of white lugugu and black lugugu in the STRUCTURE groups was corroborated by a pairwise Rst value of 0.07 (Table 3), and an average of 2.54 migrating accessions (Nm) between the two (Table 4). PCA (Figure Apx2) showed a separation of varying degrees between the less common landraces (except for roma) and the common landraces. In addition, pairwise Rst values (Table 3) showed that the less common landraces (limondigua, sandala, wawa, magaje, roma) were more similar (Rst=0.13-0.48) to the common landraces than they were to each other (Rst=0.32-1.00), although many of the Rst values were not significant, possibly due to small sample sizes of the less

common landraces (Table 3). When the most diverging landraces were removed (limondigua, wawa, magaje and sandala; Figure Apx2), a PCO (Figure 9) showed that accessions sampled from the same household had an affinity to one another, a result also supported by the STRUCTURE groups (Figure 8A). This was especially evident in Household 2, which differed markedly from the other households (Figure 8A, Figure 9). Household 4 and 5 on the other hand, had a high degree of genetic overlap also seen by a pairwise Rst value of 0.01 (Table 5).

Table 3 **Pairwise Rst values of 52 accessions of cultivated sorghum based on 17 microsatellites, consisting of eight landraces (BL=black lugugu, WL=white lugugu, R=roma, N=namata, Mg=magaje, W=wawa, Li=limondigua, Sn=sandala), collected from five households in Hombolo, Tanzania. N=sample size, L=landrace (p values ≤ 0.05 are written in bold).**

N		**16**	**19**	**7**	**4**	**3**	**1**	**1**
	L	**BL**	**WL**	**R**	**N**	**Mg**	**W**	**Li**
16	**BL**	0						
19	**WL**	**0.07**	**0**					
7	**R**	0.13	0.18	0				
4	**N**	**0.16**	**0.25**	**0.54**	0			
3	**Mg**	0.19	0.28	**0.56**	**0.19**	0		
1	**W**	0.36	0.48	0.74	0.76	0.71	0	
1	**Li**	0.40	0.45	0.77	0.73	0.58	0.99	0
1	**Sn**	0.24	0.26	0.32	0.49	0.42	0.97	1



A) **B)**

Figure 8A) Plot of STRUCTURE results (K=8) based on 17 microsatellites, for 52 cultivated sorghum accessions, collected from five households in Hombolo, Tanzania. Accessions are named according to landrace, with eight landraces represented (black lugugu, white lugugu, limondigua, wawa, magaje, roma, sandala and namata). The number preceding the landrace name refers to the household from which it was collected. The groups (K) are represented by different colors. The segmentation of the horizontal pillars shows with what percentage an accessions is placed within which groups. B) Plot of the output from STRUCTURE-SUM including a summary, for K=1-9, of the logarithmic probability (ln P(D)), the similarity coefficient and delta K.

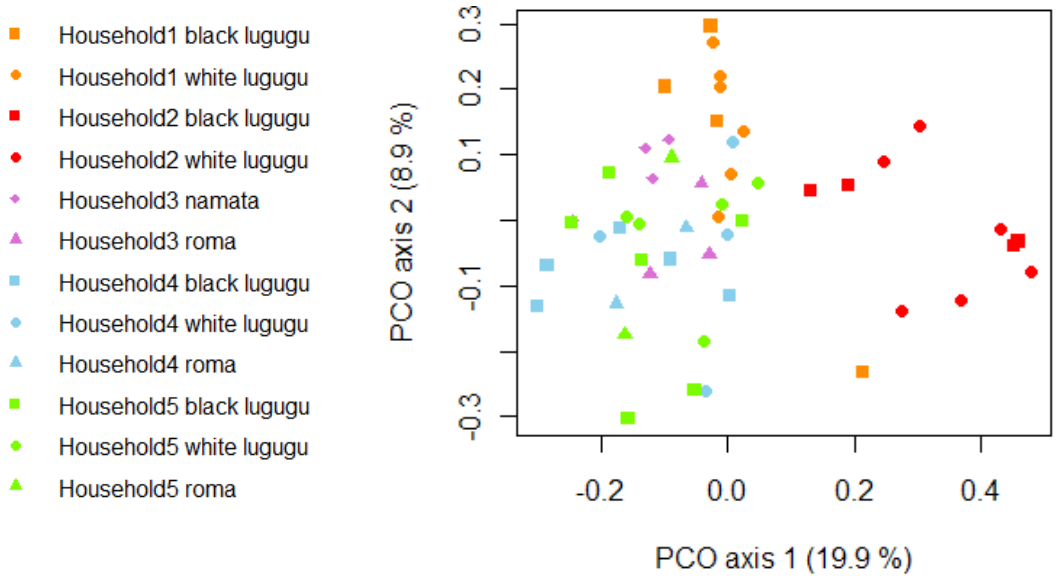


Figure 9 **Principal Coordinate Analysis (PCO) based on 17 microsatellite loci, of 52 cultivated sorghum accessions, representing four landraces (white lugugu, black lugugu, roma and namata; indicated by different symbols) sampled from five households (indicated by different colors) in Hombolo, Tanzania. The most divergent landraces (wawa, sandala, limondigua and magaje) were not included, in order to increase the resolution of the remaining accessions.**

Table 4 **Mean allelic richness (Rs), mean private allelic richness (PRs), mean expected heterozygosity (He) and mean observed heterozygosity (Ho) across 17 microsatellite loci, for 49 cultivated sorghum accessions representing five landraces (black lugugu, white lugugu, namata, roma, magaje) and 26 'wild' sorghum accessions collected from five households in Hombolo, Tanzania. Tanzania- refers to 42 cultivated sorghum accessions from throughout Tanzania and Africa- refers to 41 cultivated sorghum from throughout Africa. N- refers to the number of accessions. Nm- refers to the number of migrating accessions between two groups (Slatkin 1985); shown in brackets. Mean Nm- is the mean number of migrating accessions across all the landrace groups. Significance values using a Wilcoxon signed rank test, for comparisons of different means of He, Rs and PRS values are shown in Table Apx7.**

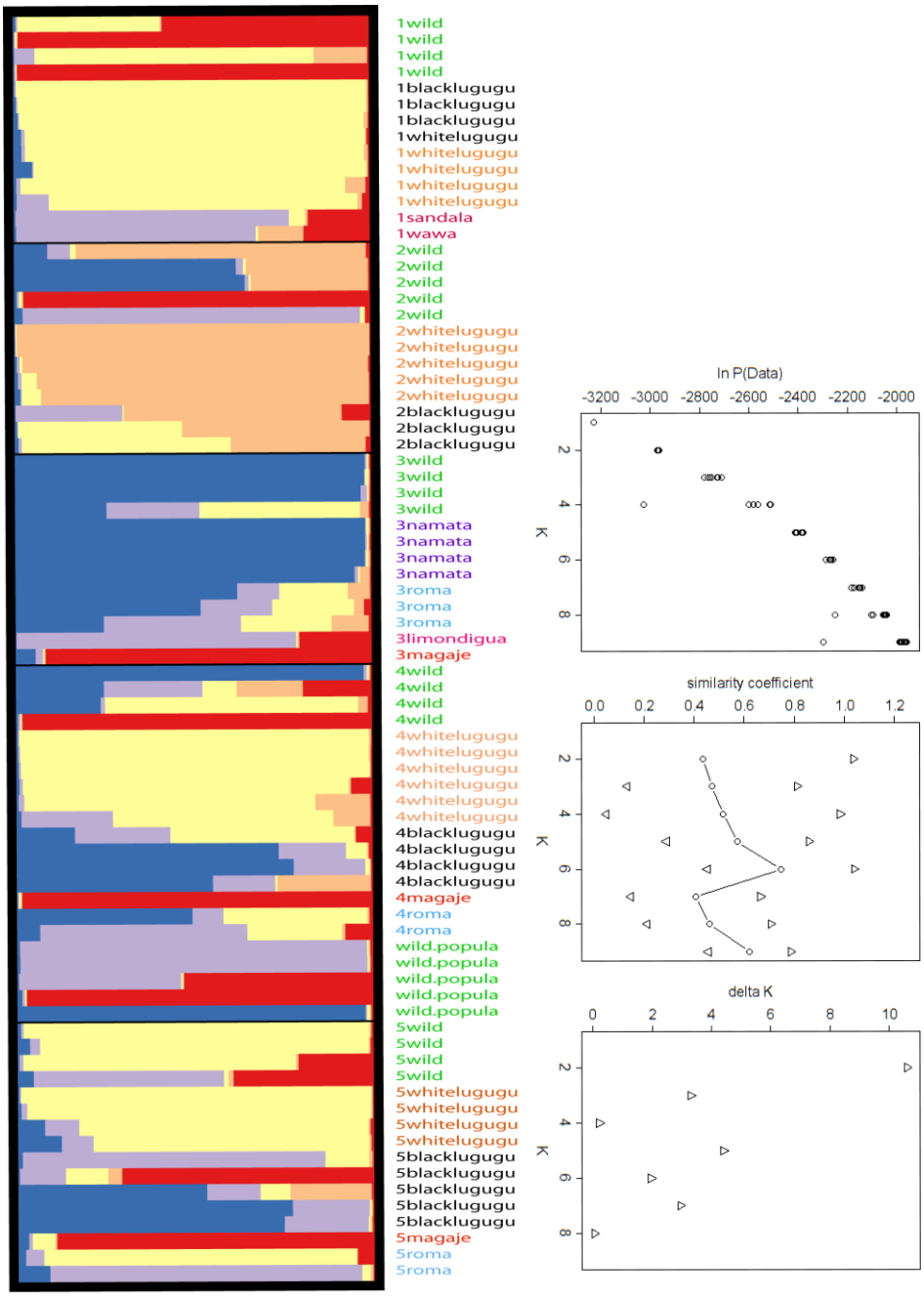
Group	Rs	PRs	He	Ho	N	Nm
black lugugu(BL)	2.09	0.40	0.49	0.23	16	**Nm(BL,WL)=2.54**
white lugugu(WL)	2.00	0.40	0.48	0.11	19	**Nm(BL,N)=0.49**
namata(N)	1.54	0.28	0.29	0.23	4	**Nm(BL,R)=1.05**
roma(R)	1.87	0.27	0.39	0.20	7	**Nm(BL,Mg)=0.30**
magaje(Mg)	1.66	0.58	0.32	0.24	3	**Nm(WL,N)=0.20**
landraces (L)	6.17	0.47	0.53	0.15	49	**Nm(WL,R)=0.63**
wild (W)	4.68	0.91	0.54	0.14	26	**Nm(WL,Mg)=0.17**
Tanzania	8.30	1.32	0.66	0.01	42	**Nm(Mg,R)=0.20**
Africa	9.60	3.33	0.72	0.01	41	**Nm(R,N)=0.40**
						Nm(L,W)=3.93
						Mean Nm =0.65

Table 5 **Pairwise Rst values based on 17 microsatellite markers, of 52 cultivated and 26 'wild' sorghum accessions collected from five households, as well as a locality of 'wild' sorghum in Hombolo, Tanzania. HH1=household 1, HH2=household 2, HH3=household 3, HH4=household 4, HH5=household 5 and Wp= 'wild' population. All the pairwise Rst values were significant ($p \leq 0.05$) using 110 permutations.**

	HH1	HH2	HH3	HH4	HH5
HH1	0				
HH2	0.18641	0			
HH3	0.20634	0.25340	0		
HH4	0.12197	0.18367	0.12735	0	
HH5	0.12466	0.18280	0.10952	0.01135	0
Wp	0.33657	0.29209	0.26787	0.21331	0.21906

Hombolo ('wild' and landraces)

For 52 cultivated and 26 'wild' sorghum accessions, a lower percentage (33%) of the locus pairs in 'wild' sorghum were in a state of linkage disequilibrium, compared with the cultivated accessions (38%; Table Apx6). A substantial amount of admixture between 'wild' and cultivated accessions were shown by AMOVA (Fst=0.03), a high number of migrating accessions (Nm=3.93), PCA (Figure Apx3) and no significant ($p \geq 0.05$) difference in allelic richness (Rs), private allelic richness (PRs) or observed heterozygosity (Ho) (Table Apx7). Five STRUCTURE groups (Figure 10A) were chosen based on the output from STRUCTURE-SUM. As the optimal K was not evident by the ln P(D) plot, the delta K plot was used to choose K (Figure 10B). There was a mixture of cultivated and 'wild' sorghum in all groups; one to three 'wild' accessions always joined the predominant group of the household from which they were collected. The remaining 'wild' accessions, as well as most of the accessions from the 'wild' population, joined other groups (Figure 10A, red and purple groups).The PCO (Figure 11) also showed that some 'wild' accessions, including accessions from the 'wild' population, clustered separately from the landraces along the second PCO axis.



A) B)

Figure 10 **A) Plot of the results from STRUCTURE (K=5) based on 17 microsatellite loci, for 78 sorghum accessions, including both cultivated and 'wild' sorghum collected from five households in Hombolo, Tanzania. The cultivated sorghum is named according to landrace, with eight landraces represented (white lugugu, black lugugu, magaje, namata, wawa, limondigua, roma and sandala). The plot is organized according to household affiliation and the number preceding the landrace names refer to the household from which the accession was collected. The groups (K) are represented by different colors. The segmentation of the horizontal pillars shows with what percentage an accessions is placed within which groups. B) Plot of the output from STRUCTURE-SUM including a summary, for k=1-9, of the logarithmic probability (ln P(D)), the similarity coefficient and delta K.**

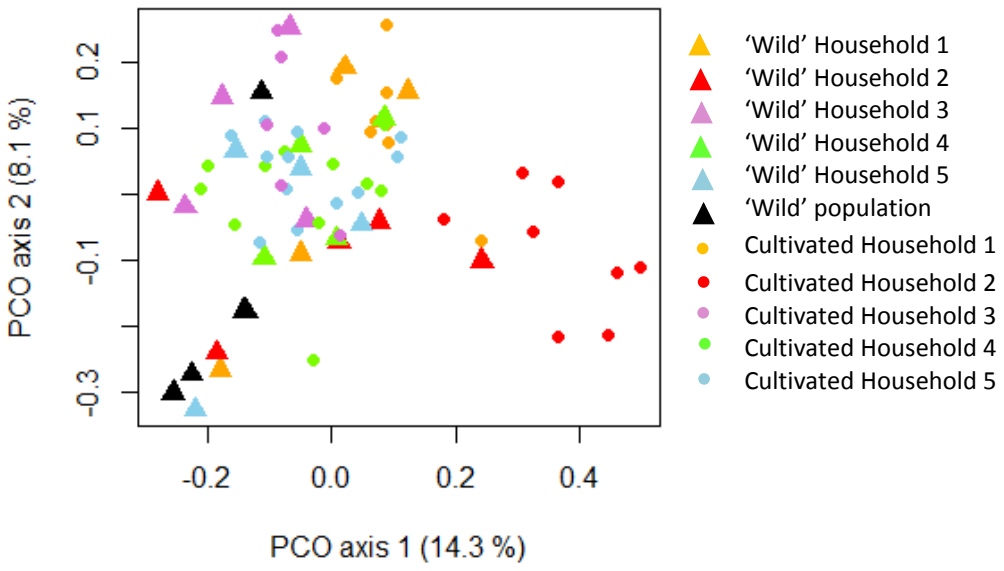


Figure 11 Principal Coordinate Analysis (PCO) based on 17 microsatellite markers, of 52 cultivated and 26 'wild' sorghum accessions from Hombolo, Tanzania. Cultivated (circle) and 'wild' (triangle) accessions were sampled from five households (indicated by different colors) and in addition, 'wild' sorghum accessions were sampled from a 'wild' sorghum population (black).

Hombolo and Tanzania

When 52 cultivated accessions from Hombolo and 42 cultivated accessions from throughout Tanzania were analyzed together, the accessions from Hombolo were somewhat separate from the remaining accessions from Tanzania as seen in the NJ tree (Figure 12), PCO (Figure 13) and STRUCTURE analysis (Figure Apx4 and Figure Apx5). Exceptions to this separation were some accessions from Dodoma, Rukwa and Mbeya, which were found amongst the accessions from Hombolo and the less common landraces (magaje, limondigua, roma (not all accessions), sandala and wawa) from Hombolo, as well as some accessions of black lugugu and white lugugu, which clustered with accessions from other regions in Tanzania. Both in the NJ tree (Figure 12) and PCO (Figure 13) most of the Hombolo accessions were placed closer to the southern Tanzanian accessions and farther away from the northern Tanzanian accessions. The NJ tree (Figure 12) corroborated most of the STRUCTURE groups (North, Northwest and Central Northwest) identified for the Tanzanian accessions, while the accessions included in the Central Southwest (blue) and Southeastern (orange) group were separated and intermingled with accessions from Hombolo in the NJ tree.

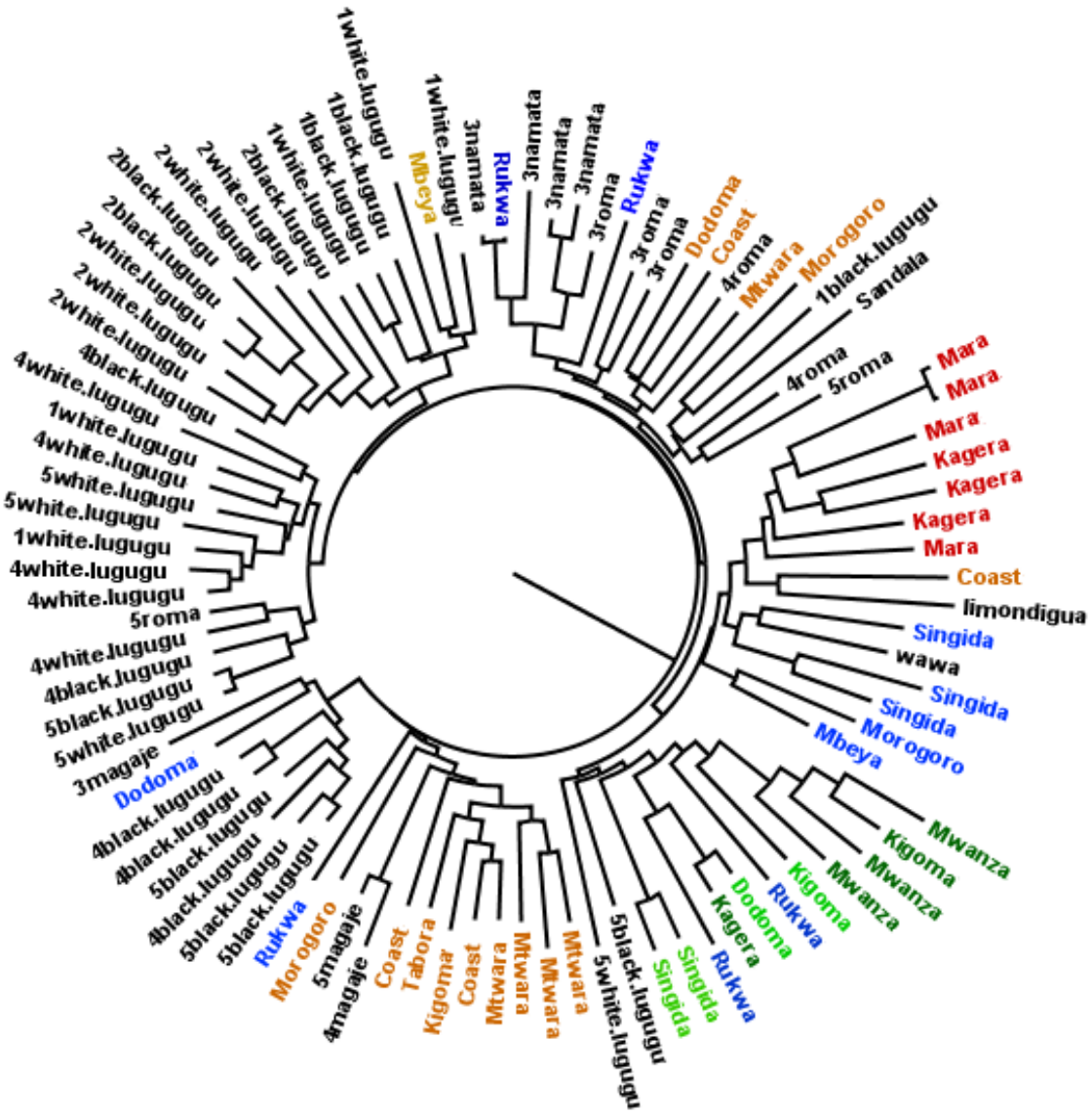


Figure 12 **Plot of a rooted neighbor joining tree of cultivated sorghum based on 17 microsatellites. The accessions represent two geographical scales: 1) local scale-52 sorghum accessions (black), collected from five households in Hombolo, Tanzania representing eight landraces (white lugugu, black lugugu, namata, roma, magaje, limondigua, sandala and wawa). The number preceding the landrace names refer to the household from which they were collected and 2) country scale- 42 sorghum accessions from throughout Tanzania (colored according to groups identified by STRUCTURE analysis (Figure 6A)).**

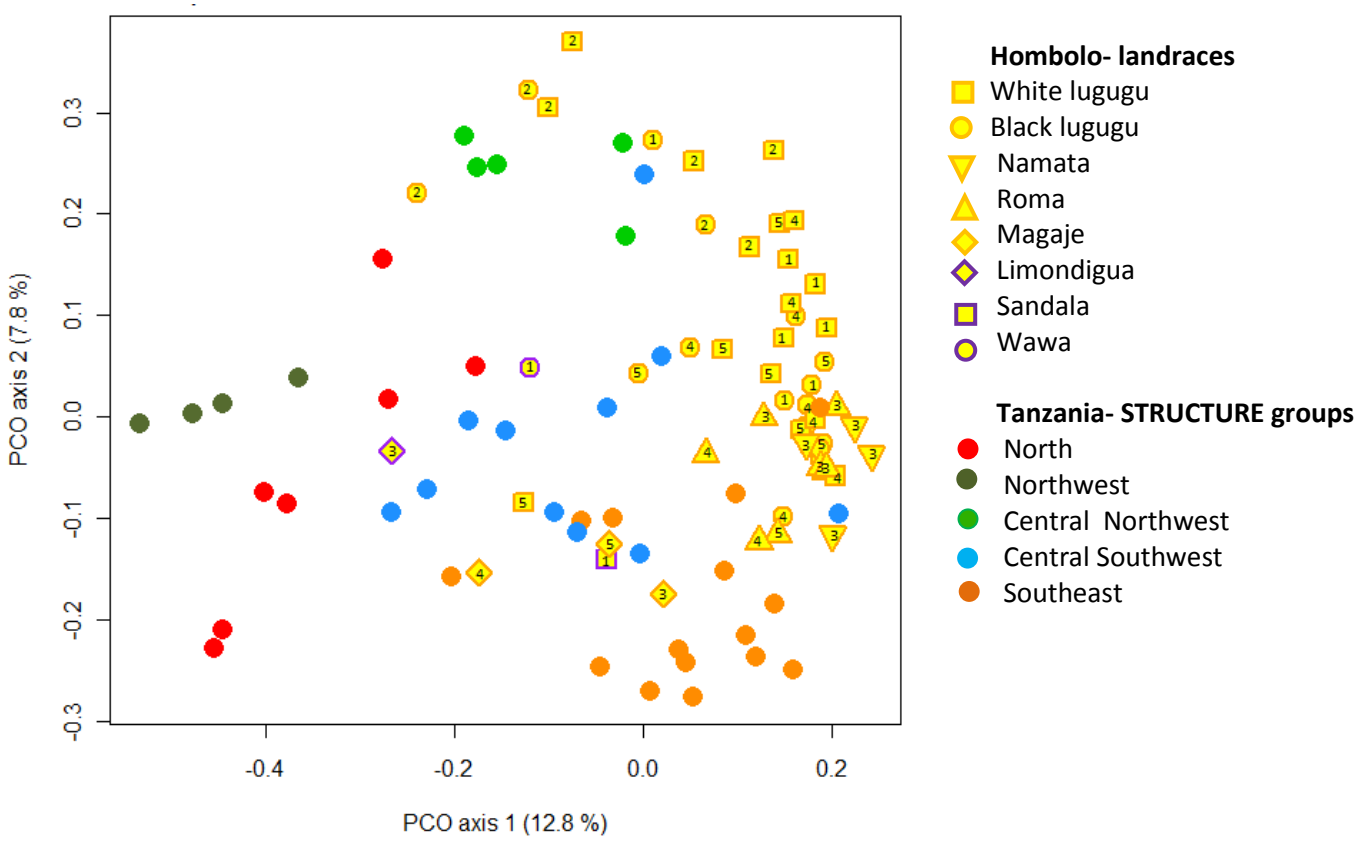


Figure 13 **Principal Coordinate Analysis (PCO) of cultivated sorghum based on 17 microsatellites. The accessions represent two different geographical scales: 1) local scale- 52 accessions, representing eight landraces, collected from five households in Hombolo, Tanzania and 2) country scale- 42 accessions from throughout Tanzania. For the accessions from Hombolo, landraces are shown by different symbols and the household from which the accessions were collected is indicated by a number (1=household1, 2=household2, 3=household3, 4=household4 and 5=household5). The accessions from throughout Tanzania are colored based on groups identified by STRUCTURE analysis (K=5). The STRUCTURE group names reflect the regions in Tanzania where the accessions, included in the groups, were collected.**

DISCUSSION

Landraces with desirable traits have been shaped through human selection over many years to meet farmers' needs. The farmers from Hombolo valued the taste and good storing abilities of their landraces, which are qualities that improved varieties in the area lacked. In addition, farmers cultivate many landraces together in one field to cover a range of culinary needs, and most importantly, as insurance for environmental fluctuations. Landraces become environmentally adapted over time, as abiotic stresses impose selection pressures in favor of tolerant landraces. Morphological diversity, however, does not always reflect genetic diversity, e.g. if the phenotypic differences are due to environmental influences. Documenting the genetic structure is a way to assess how genetic diversity is partitioned within a species, and such information may be used for *ex situ* and *in situ* conservation of genetic diversity. In this study we aimed to investigate how the genetic diversity within sorghum landraces is structured and related to geography, climate/race, landrace/grain color, if human-mediated effects or mating system have any effect on the genetic structure and lastly, how gene flow affects the genetic diversity/structure of *Sorghum bicolor*.

To what degree is the genetic diversity structured based on geography?

The genetic diversity of sorghum was found to be structured according to geography at all three geographical scales (Africa, Tanzania and Hombolo). The importance of geographical structure increased as the scales became larger, as shown by Mantel tests; Africa ($R=0.36$, $p\leq0.05$), Tanzania ($R=0.34$,$p\leq0.05$) and Hombolo ($R=0.22$, $p\leq0.05$). The African accessions clustered into an eastern, western, northeastern and southern group, with some striking exceptions. The accessions from Sudan were found in all four groups and some accessions from South Africa, India, Zimbabwe and Somalia were included in the northeastern group. These observations are compatible with the postulated origin of sorghum in northeastern Africa. Representing the center of origin of sorghum domestication, the Sudan accessions more or less displayed the diversity found elsewhere in Africa, and accessions from other parts of Africa and India may cluster with the northeastern group, because this group represents the original gene pool. Two other studies also found geographical structuring of sorghum diversity across Africa, especially a separation between southern African accessions and the rest (Deu, et al. 2006; Morris, et al. 2012). High genetic diversity amongst accessions from Sudan was also attributed to the location of Sudan in

sorghum's center of diversity in a previous study (Assar, et al. 2005). Furthermore the study by Deu, et al. (2006) also found some South African accessions clustering with northeastern African accessions and drew the same conclusions as we do. However, a more comprehensive study is needed to consolidate this theory and at present it cannot be excluded that the these observations are not due to long distance seed exchange.

The sorghum accessions from Tanzania largely clustered into a northern, northwestern, central northwestern, central southwestern and southeastern group. Previous studies of sorghum landraces at a country scale in Niger, Sudan, Kenya, Mali, Guinea, Uganda and Zambia have also found genetic diversity to be structured geographically (Deu, et al. 2008; Mbeyagala, et al. 2012; Mutegi, et al. 2011; Ng'uni, et al. 2011; Rabbi, et al. 2010; Sagnard, et al. 2011). A study in Kenya and Sudan (Rabbi, et al. 2010) and one in Kenya (Mutegi, et al. 2010) did not find any geographical structure. The explanation put forward by the authors, was that farmers in several regions of Kenya use improved varieties acquired from the formal seed sector, which has a homogenizing effect on the diversity.

A Mantel test revealed significant geographical structuring of genetic diversity for cultivated sorghum at the local scale in Hombolo, even though the distances between collection sites (households) were quite small, ranging from 150 m to 1.6 km apart. However, pairwise Fst values between the household accessions showed some inconsistencies regarding the geographical structure. Household 1 and household 2 were only located 150 m apart and still quite differentiated (Fst=0.18), while household 4 and household 5, which were 400 m apart, were genetically very similar (Fst=0.01). Together with the discovery that the farmers obtain their seeds from the neighbors, these results suggest that farmers' seed exchanging patterns are the primary force impacting the genetic structure of sorghum at a local scale. However, where seed exchanging is not present, genetic drift will act to differentiate fields of sorghum.

To what degree is the genetic diversity structured based on climate/race?

There was some genetic structuring according to race, temperature and precipitation, although this was relatively weak compared with the geographical structuring. Guinea was mainly found in western Africa, kafir in southern Africa, durra in northeastern Africa, caudatum in east/northeastern Africa and bicolor in west/northeastern Africa. This is consistent with their known distributions, reflecting agro-climatic adaptations. Guinea and caudatum are often

cultivated in tropical savannas, durra and caudatum are often cultivated in hot semi-arid climates and kafir is cultivated in temperate/sub-tropical climates (Harlan 1976). Structuring of genetic diversity according to race was also observed by Deu, et al. (2006) based on RFLP (restriction fragment length polymorphism) probes and by Morris, et al. (2012) based on genome-wide SNP (single nucleotide polymorphism) data. The general weak associations between climatic factors and genetic diversity of sorghum (which have also been observed in other microsatellite studies; Barro-Kondombo, et al. 2010; Deu, et al. 2008; Sagnard, et al. 2011) may be related to the (presumably) adaptively neutral nature of microsatellites.

To what degree is the genetic diversity structured based on landrace/grain color?

The genetic analysis of eight sorghum landraces collected in Hombolo, Tanzania showed that, on most occasions, the landrace names reflected the underlying genetic differentiation, as done for namata, roma, magaje, wawa, limondigua, and sandala. However, there was substantial genetic overlap between the two most common landraces in Hombolo (white lugugu and black lugugu). White lugugu and black lugugu have different colored glumes (Figure 4); most likely the main feature explaining their different names. The differentiation between STRUCTURE groups (Fst=0.33) was higher than the differentiation found between landraces (Fst=0.14), which is most likely due to the high genetic similarity between white lugugu and black lugugu. The level of differentiation between STRUCTURE identified groups was similar to landrace differentiation in a village in Cameroon (Fst=0.35, N=293) across 14 microsatellites (Barnaud, et al. 2007).

Underlying genetic differentiation between differently named landraces was also found in Eritrea, Somalia and Cameroon (Ghebru, et al. 2002; Manzelli, et al. 2007; Soler, et al. 2012), but not in Zimbabwe (Chakauya, et al. 2006). In the last mentioned study landraces were compared between different regions. It was discovered that landrace names, used in different villages, were not genetically alike suggesting that landrace names may be more useful as indicators of genetic differentiation within a single village, but not as much between villages. Chakauya, et al. (2006) attributed this to the obscure origin of landrace names. Other studies, however, in Ethiopia and Somalia, found correlations between morphological characters and landrace names, with grain color and glume color as important distinguishing features between landraces. Grain color was especially important as it distinguished between sweet stemmed sorghums and grain sorghums (Manzelli, et al. 2005; Teshome, et al. 1997).

In our study the differentiation according to grain color became weaker as the geographical scale became larger; Africa (Fst=-0.01), Tanzania (Fst=0.12) and Hombolo (Fst=0.19). This shows that the closer landraces are to each other geographically, the more important grain color becomes as a genetically distinguishing feature. This may also explain the lack of genetic differentiation between landraces in different regions in Zimbabwe (Chakauya, et al. 2006).

The expected heterozygosity of cultivated sorghum found in Tanzania (He=0.66, N=51) was similar but slightly higher than that found for cultivated sorghum in Kenya (He=0.59, N=439) across 24 microsatellites (Mutegi, et al. 2011) and Niger (He =0.61, N=472) across 28 microsatellites (Deu, et al. 2008), but was much higher than that found in Burkina Faso (He=0.37, N=124) across 29 microsatellites (Barro-Kondombo, et al. 2010). The explanation put forward by Barro-Kondombo for the low level of diversity found in Burkina Faso, compared with other studies, is the low amount of sorghum landrace/racial diversity present in Burkina Faso.

In what ways do human cultivation practices and mating system influence the genetic diversity and structure of cultivated sorghum?

The noticeable geographical structure of sorghum diversity across Africa and Tanzania can be explained by the cultivation of localized landraces throughout most of Africa. Farmers obtain their seeds from each other or from the recycling of their own crops. In the last mentioned case they only select a small portion of their harvest to be replanted, a practice which enhances genetic differentiation between villages and regions. In addition, different races are traditionally cultivated in different areas of Africa (explained by historical human migrations, domestication events and environmental adaptations (Kimber 2000)) enhancing geographical structure. Contrastingly the cultivation of improved varieties acquired from the formal seed sector results in homogenization of genetic diversity, as seen in studies in Kenya and Sudan (Mutegi, et al. 2010; Rabbi, et al. 2010). Population differentiation is also enhanced by sorghums self-fertilizing mating strategy. The geographical structure of local landraces of maize throughout Africa and pearl millet throughout India (both outcrossing crops) is much less apparent compared to sorghum (Chowdari, et al. 1998; Westengen, et al. 2012), possibly due to the difference in mating systems.

Sorghum sampled from fields had much higher levels of heterozygosity, than that obtained from gene banks. This phenomena was also found in a study in Eritrea (Ghebru, et al. 2002) which included both *in situ* landraces and gene bank samples. A study by Adugna, et al. 2012 also found this to be the case when comparing *in situ* wild samples with gene bank wild samples. This is most likely due to the way material is managed by gene banks. In order to retain freshness, samples are periodically replanted. During this process special care is taken to avoid cross fertilization in order to maintain the 'genetic purity' of the samples.

Does gene flow occur between landrace and between landraces and wild/weedy sorghum?

Gene flow was detected between sorghum landraces in Hombolo based on overall low pairwise Rst values and genetic overlap (as seen in ordination analyses). The less common landraces were genetically more similar to the common landraces (that they were growing together with), than the less common landraces were to each other. Despite the occurrence of gene flow (promoted by the mixing of landraces in a single field), the landraces mostly remain genetically distinct. This can be partly attributed to the fact that farmers often choose grains (for replanting) which are stereotypical of a landrace, and partly to the presence of pollen competition between different landraces (Barnaud, et al. 2008; Muraya, et al. 2011a).

Gene flow was also detected between cultivated and 'wild' sorghum plants, evident by the number of migrating accessions between the two genetic pools (Nm=3.93), and the overlapping genetic structure. Low genetic separation was also found between 'wild' and cultivated gene pools in Cameroon, Kenya, Mali, Guinea and Ethiopia (Adugna, et al. 2012; Barnaud, et al. 2009; Mutegi, et al. 2011; Sagnard, et al. 2011). A study analyzing both true wild accessions (subsp. *verticilliflorum*) and weedy accessions of sorghum (subsp. *drummondii*) found that wild accessions formed their own group, while weedy accessions clustered with landrace accessions (Casa, et al. 2005). The 'wild' sorghum collected in Hombolo shared genetic diversity levels and grouped together with the landraces, suggesting that they are weedy sorghums (subsp. *drummondii*) rather than true wild sorghum. A similar conclusion was drawn in a study in Mali by Sagnard, et al. (2011), who also found a similar genetic overlap between 'wild' and cultivated accessions. This is further supported by the morphological similarity of our 'wild' samples with pictures of weedy sorghum in the study by Okeno, et al. (2012). Studies have shown that the

morphology of wild sorghum is greatly affected by its proximity to cultivated sorghum fields, where more often than not, the 'wild' sorghum found growing within a couple of kilometers of cultivated fields are crop-wild hybrids (Okeno, et al. 2012; Tesso, et al. 2008). Another possibility for the origin of these 'wild' plants is de-domestication (endoferality; a result of crop-crop hybridization) of crop plants, although this has not been recorded before within *Sorghum bicolor* (Ellstrand, et al. 2010; Warwick and Stewart 2005). Yet the endoferal origin of weedy sorghum remains a plausible possibility due to the close genetic proximity found between cultivated and weedy sorghum.

Our results show much higher rates of crop-'wild' gene flow (Nm=3.93, Fst=0.03) than crop-crop gene flow (mean Nm=0.65, Fst=0.14), despite the mixture of landraces in the fields. The higher crop-'wild' gene flow, compared with the crop-crop gene flow, could be explained by higher outcrossing rates in 'wild' compared with cultivated sorghum. This could be due to the differences' in panicle density, as shown in the study by Dje, et al. (2004). Wild (or weedy) sorghum has looser panicles compared with cultivated sorghum, facilitating higher crop-wild gene flow compared with the crop-crop gene flow. Higher outcrossing rates may also explain the lower amount of linkage disequilibrium found in our 'wild' accessions, compared with the landraces, and the lack of geographical structuring of genetic diversity among the 'wild' accessions.

Implications for conservation

This study shows that the genetic diversity of sorghum landraces in Africa is geographically structured, reflecting the diversity of landraces present across the African continent, a country (Tanzania) and a single village (Hombolo). Traditional landrace cultivation practices maintains a considerable diversity in farming systems (Bezançon, et al. 2009; Jarvis, et al. 2008) which is absent under industrialized cultivation practices (which is dominant in most developed countries). The existent diversity of sorghum throughout Africa is important to conserve for its present role in food security and for the role it may play for the food security needs in the future (Plucknett, et al. 1987). Most of the sorghum diversity present in Africa has not yet been harnessed by breeding programs (Dillon, et al. 2007) and could contribute significantly to sorghum production in Africa. Breeding programs usually utilize the diversity present within gene banks for crop improvement. It has, however, been shown that major gene bank collections

only represent a small share of the diversity of landraces found *in situ*, especially for countries in Sub-Saharan Africa, where sorghum is adapted to climates analogous of those which may befall other areas in the future (Burke, et al. 2009). The study by Burke, et al. (2009) did not, however, evaluate gene bank accessions present in ICRISAT, or national gene banks such as NPGRC, therefore more research may be needed to properly investigate this. Landraces on farm are acknowledged as the main source of genetic diversity for gene banks and breeding programs, yet many studies have shown genetic erosion due to the introduction of improved varieties (Frankel 1970; Newton, et al. 2009; Shewayrga, et al. 2008; Thomas, et al. 2011). It has been suggested, however, that only landraces which are not used for specific reasons are subject to genetic erosion, while those which are selected by farmers for certain desirable traits (and have been for years) are likely to survive on farm alongside improved varieties (Berg 2009).

REFERENCES

Adugna A, Snow A, Sweeney P, Bekele E, Mutegi E 2012. Population genetic structure of in situ wild *Sorghum bicolor* in its Ethiopian center of origin based on SSR markers. Genetic Resources and Crop Evolution 1: 1-16.

Ahmed M, Sanders JH, Nell W 2000. New sorghum and millet cultivar introduction in Sub-Saharan Africa: impacts and research agenda. Agricultural Systems 64: 55-65.

Arriola PE, Ellstrand NC 1996. Crop-to-weed gene flow in the genus *Sorghum* (Poaceae): spontaneous interspecific hybridization between johnsongrass, *Sorghum halepense*, and crop sorghum, *S. bicolor*. American Journal of Botany 1: 1153-1159.

Arriola PE, Ellstrand NC 1997. Fitness of interspecific hybrids in the genus Sorghum: persistence of crop genes in wild populations. Ecological Applications 7: 512-518.

Assar A, Uptmoor R, Abdelmula A, Salih M, Ordon F, Friedt W 2005. Genetic variation in sorghum germplasm from Sudan, ICRISAT, and USA assessed by simple sequence repeats (SSRs). Crop Science 45: 1636-1644.

Ball CR 1930. The history of American wheat improvement. Agricultural History 4: 48-71.

Barakat H, Fahmy A. 1999. Wild grasses as 'Neolithic' food resources in the eastern Sahara. In: Veen Mvd, editor. The exploitation of plant resources in ancient Africa. New York, USA: Plenum publishers.

Barnaud A, Deu M, Garine E, Chantereau J, Bolteu J, Koida EO, McKey D, Joly HI 2009. A weed-crop complex in sorghum: the dynamics of genetic diversity in a traditional farming system. American Journal of Botany 96: 1869-1879.

Barnaud A, Deu M, Garine E, McKey D, Joly HI 2007. Local genetic diversity of sorghum in a village in northern Cameroon: structure and dynamics of landraces. Theoretical and Applied Genetics 114: 237-248.

Barnaud A, Trigueros G, McKey D, Joly HI 2008. High outcrossing rates in fields with mixed sorghum landraces: how are landraces maintained? Heredity 101: 445-452.

Barro-Kondombo C, Sagnard F, Chantereau J, Deu M, vom Brocke K, Durand P, Goze E, Zongo JD 2010. Genetic structure among sorghum landraces as revealed by morphological variation and microsatellite markers in three agroclimatic regions of Burkina Faso. Theoretical and Applied Genetics 120: 1511-1523.

Berg T 2009. Landraces and folk varieties: a conceptual reappraisal of terminology. Euphytica 166: 423-430.

Bezançon G, Pham JL, Deu M, Vigouroux Y, Sagnard F, Mariac C, Kapran I, Mamadou A, Gérard B, Ndjeunga J 2009. Changes in the diversity and geographic distribution of cultivated millet (*Pennisetum glaucum* (L.) R. Br.) and sorghum (*Sorghum bicolor* (L.) Moench) varieties in Niger between 1976 and 2003. Genetic Resources and Crop Evolution 56: 223-236.

Bhattramakki D, Dong J, Chhabra AK, Hart GE 2000. An integrated SSR and RFLP linkage map of *Sorghum bicolor* (L.) Moench. Genome 43: 988-1002.

Billot C, Rivallan R, Sall MN, Fonceka D, Deu M, Glaszmann JC, Noyer JL, Rami JF, Risterucci AM, Wincker P, Ramu P, Hash CT 2012. A reference microsatellite kit to assess for genetic diversity of *Sorghum bicolor* (Poaceae). American Journal of Botany 99: 245-250.

Bonnet E, Van de Peer Y 2002. zt: a software tool for simple and partial Mantel tests. Journal of Statistical Software 7: 1-12.

Botha GM, Viljoen CD 2008. Can GM sorghum impact Africa? Trends in biotechnology 26: 64-69.
Brown S, Hopkins M, Mitchell S, Senior M, Wang T, Duncan R, Gonzalez-Candelas F, Kresovich S 1996. Multiple methods for the identification of polymorphic simple sequence repeats (SSRs) in sorghum [*Sorghum bicolor* (L.) Moench]. Theoretical and Applied Genetics 93: 190-198.

Brunken J, de Wet JMJ, Harlan J 1977. The morphology and domestication of pearl millet. Economic Botany 31: 163-174.

Burke MB, Lobell DB, Guarino L 2009. Shifts in African crop climates by 2050, and the implications for crop improvement and genetic resources conservation. Global Environmental Change 19: 317-325.

Casa A, Mitchell S, Hamblin M, Sun H, Bowers J, Paterson A, Aquadro C, Kresovich S 2005. Diversity and selection in sorghum: simultaneous analyses using simple sequence repeats. Theoretical and Applied Genetics 111: 23-30.

Cavatassi R, Lipper L, Narloch U 2011. Modern variety adoption and risk management in drought prone areas: insights from the sorghum farmers of eastern Ethiopia. Agricultural Economics 42: 279-292.

Chakauya E, Tongoona P, Matibiri E, Grum M 2006. Genetic diversity assessment of sorghum landraces in Zimbabwe using microsatellites and indigenous local names. International Journal of Botany 2: 29-35.

Cheverud J, Routman E, Jaquish C, Tardif S, Peterson G, Belfiore N, Forman L 2002. Quantitative and molecular genetic variation in captive cotton-top tamarins (*Saguinus oedipus*). Conservation Biology 8: 95-105.

Chowdari K, Davierwala A, Gupta V, Ranjekar P, Govila O 1998. Genotype identification and assessment of genetic relationships in pearl millet [*Pennisetum glaucum* (L.) R. Br] using microsatellites and RAPDs. Theoretical and Applied Genetics 97: 154-162.

Clark JD, Stemler A 1975. Early domesticated sorghum from Central Sudan. Nature 254: 588-591.
Conner AJ, Glare TR, Nap JP 2003. The release of genetically modified crops into the environment. The Plant Journal 33: 19-46.

Darwin CR. 1868. Variation of plants and animals under domestication. London, UK: William Clowes and Sons.

De Wet J 1978. Systematics and evolution of *Sorghum* sect. *Sorghum* (Gramineae). American Journal of Botany 65: 477-484.

De Wet JMJ, Harlan J 1971. The origin and domestication of *Sorghum bicolor*. Economic Botany 25: 128-135.

Deu M, Rattunde F, Chantereau J 2006. A global view of genetic diversity in cultivated sorghums using a core collection. Genome 49: 168-180.

Deu M, Sagnard F, Chantereau J, Calatayud C, Hérault D, Mariac C, Pham JL, Vigouroux Y, Kapran I, Traore PS, Mamadou A, Gerard B, Ndjeunga J, Bezançon G 2008. Niger-wide assessment of *in situ* sorghum genetic diversity with microsatellite markers. Theoretical and Applied Genetics 116: 903-913.

Dietz EJ 1983. Permutation tests for association between two distance matrices. Systematic Biology 32: 21-26.

Dillon SL, Shapter FM, Henry RJ, Cordeiro G, Izquierdo L, Lee LS 2007. Domestication to crop improvement: genetic resources for *Sorghum* and *Saccharum* (Andropogoneae). Annals of Botany 100: 975-989.

Dje Y, Heuertz M, Ater M, Lefèbvre C, Vekemans X 2004. *In situ* estimation of outcrossing rate in sorghum landraces using microsatellite markers. Euphytica 138: 205-212.

Doggett H. 1991. Sorghum history in relation to Ethiopia. In: Engels JMM, editor. Plant Genetic Resources of Ethiopia. Cambridge, UK: Press Syndicate of the University of Cambridge.

Duncan R, Dahlberg J, Spinks M. 1995. International activities in sorghum germplasm acquisition during the past thirty-five years. In. International Germplasm Transfer: Past and Present. Illinois, USA: Crop Science Society of America and American Society of Agronomy.

Ehrich D 2006. AFLPdat: a collection of R functions for convenient handling of AFLP data. Molecular Ecology Notes 6: 603-604.

Ellstrand N, Foster K 1983. Impact of population structure on the apparent outcrossing rate of grain sorghum (*Sorghum bicolor*). Theoretical and Applied Genetics 66: 323-327.

Ellstrand NC, Heredia SM, Leak-Garcia JA, Heraty JM, Burger JC, Yao L, Nohzadeh-Malakshah S, Ridley CE 2010. Crops gone wild: evolution of weeds and invasives from domesticated ancestors. Evolutionary Applications 3: 494-504.

Ellstrand NC, Hoffman CA 1990. Hybridization as an avenue of escape for engineered genes. BioScience 40: 438-442.

Ellstrand NC, Prentice HC, Hancock JF 1999. Gene flow and introgression from domesticated plants into their wild relatives. Annual Review of Ecology and Systematics 1: 539-563.

Ellstrand NC, Schierenbeck KA 2000. Hybridization as a stimulus for the evolution of invasiveness in plants? Proceedings of the National Academy of Sciences 97: 7043-7050.

Evanno G, Regnaut S, Goudet J 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14: 2611-2620.

Evenson RE, Gollin D 2003. Assessing the impact of the Green Revolution, 1960 to 2000. Science 300: 758-762.

Excoffier L, Laval G, Schneider S 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1: 47-50.

Eyre-Walker A, Gaut RL, Hilton H, Feldman DL, Gaut BS 1998. Investigation of the bottleneck leading to the domestication of maize. Proceedings of the National Academy of Sciences 95: 4441-4446.

FAO.F 1999. Post-harvest operations: Sorghum [Internet], cited December 2012 [www.fao.org].

Fattovich R, Marks AE, Mohammed-Ali A 1984. The archaeology of the Eastern Sahel, Sudan: preliminary results. African Archaeological Review 2: 173-188.

Frankel SOH 1970. Genetic conservation of plants useful to man. Biological Conservation 2: 162-168.

Funk C, Dettinger MD, Michaelsen JC, Verdin JP, Brown ME, Barlow M, Hoell A 2008. Warming of the Indian Ocean threatens eastern and southern African food security but could be mitigated by agricultural development. Proceedings of the National Academy of Sciences 105: 11081-11086.

Futuyma DJ. 2009. Evolution. Massachusetts, USA: Sinauer Associates.

Gao H, Williamson S, Bustamante CD 2007. A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. Genetics 176: 1635-1651.

Gepts P, Papa R 2003. Possible effects of (trans)gene flow from crops on the genetic diversity from landraces and wild relatives. Environmental Biosafety Research 2: 89-103.

Ghebru B, Schmidt R, Bennetzen J 2002. Genetic diversity of Eritrean sorghum landraces assessed with simple sequence repeat (SSR) markers. Theoretical and Applied Genetics 105: 229-236.

Glaubitz JC 2004. convert: A user-friendly program to reformat diploid genotypic data for commonly used population genetic software packages. Molecular Ecology Notes 4: 309-310.

Guo J, Wang Y, Song C, Zhou J, Qiu L, Huang H, Wang Y 2010. A single origin and moderate bottleneck during domestication of soybean (*Glycine max*): implications from microsatellites and nucleotide sequences. Annals of Botany 106: 505-514.

Hammer O, Harper D, Ryan P 2001. PAST: Paleontological statistics software package for education and data analysis. Paleontologica Electronica 4: 1-9.

Harlan J, De Wet J 1972. A simplified classification of cultivated sorghum. Crop Science 12: 172-176.

Harlan JR. 1976. Origins of African plant domestication. Amsterdam, Netherlands: De Gruyter Mouton.

Haudry A, Cenci A, Ravel C, Bataillon T, Brunel D, Poncet C, Hochu I, Poirier S, Santoni S, Glemin S 2007. Grinding up wheat: a massive loss of nucleotide diversity since domestication. Molecular Biology and Evolution 24: 1506-1517.

Holm LR 1969. Weeds problems in developing countries. Weed Science 17: 113-118.

House LR. 1985. A guide to sorghum breeding. Patancheru, India: ICRISAT (International Crops Research Institute for the Semi-Arid Tropics).

ICRISAT (International Crops Research Institute for the Semi-Arid Tropics). 2010. [Internet], cited January 2013 [http://www.icrisat.org/crop-sorghum.htm].

Jarvis DI, Brown AHD, Cuong PH, Collado-Panduro L, Latournerie-Moreno L, Gyawali S, Tanto T, Sawadogo M, Mar I, Sadiki M 2008. A global perspective of the richness and evenness of traditional crop-variety diversity maintained by farming communities. Proceedings of the National Academy of Sciences 105: 5326-5331.

Jarvis DI, Hodgkin T 2002. Wild relatives and crop cultivars: detecting natural introgression and farmer selection of new genetic combinations in agroecosystems. Molecular Ecology 8: 159-173.

Jombart T 2008. adegenet: an R package for the multivariate analysis of genetic markers. Bioinformatics 24: 1403-1405.

Kajale M 1977. On the botanical findings from excavations at Daimabad, a Chalcolithic site in Western Maharashtra, India. Current Science 49: 818-819.

Kalinowski ST 2005. hp-rare 1.0: a computer program for performing rarefaction on measures of allelic richness. Molecular Ecology Notes 5: 187-189.

Kamala V, Sharma H, Manohar Rao D, Varaprasad K, Bramel P 2009. Wild relatives of sorghum as sources of resistance to sorghum shoot fly, Atherigona soccata. Plant Breeding 128: 137-142.

Kamala V, Singh S, Bramel P, Rao DM 2002. Sources of resistance to downy mildew in wild and weedy sorghums. Crop Science 42: 1357-1360.

Kimber C. 2000. Origins of domesticated sorghum and its early diffusion to India and China. In: Cam Wallas Smith RAF, editor. Sorghum: Origin, history, technology, and production. New York, USA: John Wiley & Sons.

Klichowska M. 1984. Plants of the Neolithic Kadero (Central Sudan): a palaeoethnobotanical study of the plant impressions on pottery. In: Lina Krzyżaniak MK, editor. Origins and early development of food-producing cultures in north-eastern Africa. Poznafi, Poland: Poznan Archaeological Museum.

Knox J, Hess T, Daccache A, Wheeler T 2012. Climate change impacts on crop productivity in Africa and South Asia. Environmental Research Letters 7: 32-34.

Kong L, Dong J, Hart G 2000. Characteristics, linkage-map positions, and allelic differentiation of Sorghum bicolor (L.) Moench DNA simple-sequence repeats (SSRs). Theoretical and Applied Genetics 101: 438-448.

Li R, Zhang H, Zhou X, Guan Y, Yao F, Song G, Wang J, Zhang C 2010. Genetic diversity in Chinese sorghum landraces revealed by chloroplast simple sequence repeats. Genetic Resources and Crop Evolution 57: 1-15.

Li Y, Li C 1998. Genetic contribution of Chinese landraces to the development of sorghum hybrids. Euphytica 102: 47-57.

Lobell DB, Burke MB, Tebaldi C, Mastrandrea MD, Falcon WP, Naylor RL 2008. Prioritizing climate change adaptation needs for food security in 2030. Science 319: 607-610.

Makanda I, Tongoona P, Derera J, Sibiya J, Fato P 2010. Combining ability and cultivar superiority of sorghum germplasm for grain yield across tropical low-and mid-altitude environments. Field Crops Research 116: 75-85.

Mantel N 1967. The detection of disease clustering and a generalized regression approach. Cancer Research 27: 209-220.

Manzelli M, Benedettelli S, Vecchio V 2005. Agricultural biodiversity in Northwest Somalia–an assessment among selected Somali sorghum (*Sorghum bicolor* (L.) Moench) germplasm. Biodiversity and Conservation 14: 3381-3392.

Manzelli M, Pileri L, Lacerenza N, Benedettelli S, Vecchio V 2007. Genetic diversity assessment in Somali sorghum (*Sorghum bicolor* (L.) Moench) accessions using microsatellite markers. Biodiversity and Conservation 16: 1715-1730.

Maqbool SB, Devi P, Sticklen MB 2001. Biotechnology: Genetic improvement of sorghum (*Sorghum bicolor* (L.) Moench). In Vitro Cellular & Developmental Biology-Plant 37: 504-515.

Mbeyagala EK, Kiambi DD, Okori P, Edema R 2012. Molecular diversity among sorghum (*Sorghum bicolor* (L.) Moench) landraces in Uganda. International Journal of Botany 8: 85-95.

Mehra K. 1991. Pre-historic Ethiopia and India: contacts through sorghum and millet genetic resources. Cambridge, UK: Cambridge University Press.

Mercader J 2009. Mozambican grass seed consumption during the Middle Stone Age. Science 326: 1680-1683.

Monaghan N 1979. The biology of Johnson grass (*Sorghum halepense*). Weed Research 19: 261-267.

Morris GP, Ramu P, Deshpande SP, Hash CT, Shah T, Upadhyaya HD, Riera-Lizarazu O, Brown PJ, Acharya CB, Mitchell SE 2012. Population genomic and genome-wide association studies of agroclimatic traits in sorghum. Proceedings of the National Academy of Sciences 110: 453-458.

Muraya M, Geiger H, Villiers S, Sagnard F, Kanyenji B, Kiambi D, Parzies H 2011a. Investigation of pollen competition between wild and cultivated sorghums (*Sorghum bicolor* (L.) Moench) using simple sequence repeats markers. Euphytica 178: 393-401.

Muraya MM, de Villiers S, Parzies HK, Mutegi E, Sagnard F, Kanyenji BM, Kiambi D, Geiger HH 2011b. Genetic structure and diversity of wild sorghum populations (*Sorghum* spp.) from different eco-geographical regions of Kenya. Theoretical and Applied Genetics 123: 571-583.

Murdock GP 1960. Staple subsistence crops of Africa. Geographical Review 50: 523-540.

Mutegi E, Sagnard F, Labuschagne M, Herselman L, Semagn K, Deu M, de Villiers S, Kanyenji BM, Mwongera CN, Traore PCS, Kiambi D 2012. Local scale patterns of gene flow and genetic diversity in a crop-wild-weedy complex of sorghum (*Sorghum bicolor* (L.) Moench) under traditional agricultural field conditions in Kenya. Conservation Genetics 13: 1059-1071.

Mutegi E, Sagnard F, Muraya M, Kanyenji B, Rono B, Mwongera C, Marangu C, Kamau J, Parzies H, de Villiers S 2010. Ecogeographical distribution of wild, weedy and cultivated *Sorghum bicolor* (L.) Moench in Kenya: implications for conservation and crop-to-wild gene flow. Genetic Resources and Crop Evolution 57: 243-253.

Mutegi E, Sagnard F, Semagn K, Deu M, Muraya M, Kanyenji B, de Villiers S, Kiambi D, Herselman L, Labuschagne M 2011. Genetic structure and relationships within and between cultivated and wild sorghum (*Sorghum bicolor* (L.) Moench) in Kenya as revealed by microsatellite markers. Theoretical and Applied Genetics 122: 989-1004.

Newton A, Akar T, Baresel J, Bebeli P, Bettencourt E, Bladenopoulos K, Czembor J, Fasoula D, Katsiotis A, Koutis K 2009. Cereal landraces for sustainable agriculture. A review. Agronomy for Sustainable Development 30: 237-269.

Ng'uni D, Geleta M, Bryngelsson T 2011. Genetic diversity in sorghum (*Sorghum bicolor* (L.) Moench) accessions of Zambia as revealed by simple sequence repeats (SSR). Hereditas 148: 52-62.

Nordborg M, Hu TT, Ishino Y, Jhaveri J, Toomajian C, Zheng H, Bakker E, Calabrese P, Gladstone J, Goyal R 2005. The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biology 3: 1289-1299.

Okeno JA, Mutegi E, de Villiers S, Wolt JD, Misra MK 2012. Morphological variation in the wild-weedy complex of *Sorghum bicolor* in situ in western Kenya: preliminary evidence of crop-wild gene flow? International Journal of Plant Sciences 173: 507-515.

Parry ML, Canziani OF, Palutikof JP, van der Linden PJ, Hanson CE. 2007. IPCC, 2007: climate change 2007: impacts, adaptation and vulnerability. Contribution of working group II to the fourth assessment report of the intergovernmental panel on climate change. Cambridge, UK: Cambridge University Press.

Pielou EC. 1984. The interpretation of ecological data: a primer on classification and ordination. California, USA: John Wiley and Sons.

Plucknett DL, Smith NJH, Williams J, Anishetty N. 1987. Gene banks and the world's food. Princeton, USA: Princeton University Press.

Pritchard JK, Stephens M, Donnelly P 2000. Inference of population structure using multilocus genotype data. Genetics 155: 945-959.

Rabbi IY, Geiger HH, Haussmann BIG, Kiambi D, Folkertsma R, Parzies HK 2010. Impact of farmers' practices and seed systems on the genetic structure of common sorghum varieties in Kenya and Sudan. Plant Genetic Resources 8: 116-126.

Rai K, Murty D, Andrews D, Bramel-Cox P 1999. Genetic enhancement of pearl millet and sorghum for the semi-arid tropics of Asia and Africa. Genome 42: 617-628.

Rambaut A .2008. Figtree version 1.2 software [Internet], cited November 2012 [http://tree.bio.ed.ac.uk/software/figtree/].

Rich PJ, Grenier C, Ejeta G 2004. Striga resistance in the wild relatives of sorghum. Crop Science 44: 2221-2229.

Ringo J 2009. Breeding and selection of sorghum (*Sorghum bicolor* (L.) Moench) for adaptation to acid soils in Tanzania. [Masters thesis], Moi University, Kenya: department of biotechnology.

Rosenberg NA 2003. DISTRUCT: a program for the graphical display of population structure. Molecular Ecology Notes 4: 137-138.

Rosenzweig C, Parry ML 1994. Potential impact of climate change on world food supply. Nature 367: 133-138.

Sagnard F, Deu M, Dembele D, Leblois R, Toure L, Diakite M, Calatayud C, Vaksmann M, Bouchet S, Malle Y, Togola S, Traore PCS 2011. Genetic diversity, structure, gene flow and evolutionary relationships within the *Sorghum bicolor* wild-weedy-crop complex in a western African region. Theoretical and Applied Genetics 123: 1231-1246.

Saitou N, Nei M 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4: 406-425.

Schlenker W, Lobell DB 2010. Robust negative impacts of climate change on African agriculture. Environmental Research Letters 5: 10-14.

Schloss S, Mitchell S, White G, Kukatla R, Bowers J, Paterson A, Kresovich S 2002. Characterization of RFLP probe sequences for gene discovery and SSR development in *Sorghum bicolor* (L.) Moench. Theoretical and Applied Genetics 105: 912-920.

Schuelke M 2000. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18: 233-234.

Seboka B, van Hintum T 2006. The dynamics of on-farm management of sorghum in Ethiopia: Implication for the conservation and improvement of plant genetic resources. Genetic Resources and Crop Evolution 53: 1385-1403.

Shewayrga H, Jordan D, Godwin I 2008. Genetic erosion and changes in distribution of sorghum (*Sorghum bicolor* (L.) Moench) landraces in north-eastern Ethiopia. Plant Genetic Resources 6: 155-158.

Slatkin M 1985. Gene flow in natural populations. Annual Review of Ecology and Systematics 1: 393-430.

Soler C, Saidou AA, Hamadou TVC, Pautasso M, Wencelius J, Joly HHI 2012. Correspondence between genetic structure and farmers' taxonomy–a case study from dry-season sorghum landraces in northern Cameroon. Plant Genetic Resources 1: 1-14.

Stewart CN, Halfhill MD, Warwick SI 2003. Transgene introgression from genetically modified crops to their wild relatives. Nature Reviews Genetics 4: 806-817.

Storfer A 1996. Quantitative genetics: a promising approach for the assessment of genetic variation in endangered species. Trends in Ecology & Evolution 11: 343-348.

Swaminathan MS 2006. An evergreen revolution. Crop Science 46: 2293-2303.

Takebayashi N, Morrell PL 2001. Is self-fertilization an evolutionary dead end? Revisiting an old hypothesis with genetic theories and a macroevolutionary approach. American Journal of Botany 88: 1143-1150.

Taramino G, Tarchini R, Ferrario S, Lee M, Pe' M 1997. Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theoretical and Applied Genetics 95: 66-72.

R Development Core Team 2010. R version 2.11. R Project for Statistical Computing Vienna, Austria.

Teshome A, Baum B, Fahrig L, Torrance J, Arnason T, Lambert J 1997. Sorghum [*Sorghum bicolor* (L.) Moench] landrace variation and classification in north Shewa and south Welo, Ethiopia. Euphytica 97: 255-263.

Tesso T, Kapran I, Grenier C, Snow A, Sweeney P, Pedersen J, Marx D, Bothma G, Ejeta G 2008. The potential for crop-to-wild gene flow in sorghum in Ethiopia and Niger: a geographic survey. Crop Science 48: 1425-1431.

Thomas M, Dawson JC, Goldringer I, Bonneuil C 2011. Seed exchanges, a key to analyze crop diversity dynamics in farmer-led on-farm conservation. Genetic Resources and Crop Evolution 58: 321-338.

UNDP. 2012. Africa Human Development Report 2012 Towards a Food Secure Future, United Nations Development Programme. New York, USA: United Nations Publications.

Vietmeyer N. 1996. Lost crops of Africa: v. 1 Grains. Washington, USA: National Academic Press.

Vigouroux Y, Barnaud A, Scarcelli N, Thuillet AC 2011. Biodiversity, evolution and adaptation of cultivated crops. Comptes Rendus Biologies 334: 450-457.

Warwick SI, Stewart C 2005. Crops come from wild plants-How domestication, transgenes, and linkage together shape ferality. Crop Ferality and Volunteerism 1: 9-30.

Wasylikowa K, Dahlberg J. 1999. Sorghum in the economy of the early Neolithic nomadic tribes at Nabta Playa, southern Egypt. In: Veen Mvd, editor. The exploitation of plant resources in ancient Africa New York, USA: Plenum Publishers.

Westengen OT, Berg PR, Kent MP, Brysting AK 2012. Spatial structure and climatic adaptation in African maize revealed by surveying SNP diversity in relation to global breeding and landrace panels. PloS One 7: 55-58.

Wiersema JH, Dahlberg J 2007. The nomenclature of *Sorghum bicolor* (L.) Moench (Gramineae). Taxon 1: 941-946.

Wilcoxon F 1945. Individual comparisons by ranking methods. Biometrics Bulletin 1: 80-83.

Zhu Q, Zheng X, Luo J, Gaut BS, Ge S 2007. Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. Molecular Biology and Evolution 24: 875-888.

APPENDIX

Tables

Table Apx1 Sorghum accessions from which DNA was extracted. The accessions marked with * were excluded from the final analysis. The table includes the accession number (Acc.), a more descriptive name used in plots (Name), the color of the grains (grain.col) and source of the grains (Source) S.P= sampled personally, NPGRC=National Plant Genetic research Center, ICRISAT= International Crops Research Institute for the Semi-Arid Tropics. Locational information is given by Lat=latitude, Long=longitude, Coll.site=collection site, Temp=temperature, Prec=precipitation, wild.pop=wild population, Morogo=Morogoro, bl.white=bleach white and NA=data not available.

Acc.	Name	grain.col	Source	Lat	Long	Coll. site	Province	Country	Temp	Prec
H1BL1	**1black.lugugu**	**white**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1BL2	**1black.lugugu**	**white**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1BL4	**1black.lugugu**	**white**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1BL5*	1black.lugugu	white	S.P	-5.96638	35.978056	Hombolo	Dodoma	TZA	22.5	750
H1SN	**1sandala**	**white**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1W1*	1wild	green	S.P	-5.96638	35.978056	Hombolo	Dodoma	TZA	22.5	750
H1W2	**1wild**	**NA**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1W3	**1wild**	**black**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1W4	**1wild**	**green**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1W5	**1wild**	**black**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1WL1	**1white.lugugu**	**white**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1WL2	**1white.lugugu**	**white**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1WL3	**1white.lugugu**	**white**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1WL4	**1white.lugugu**	**white**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1WL5	**1white.lugugu**	**white**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H1WW	**1wawa**	**red**	**S.P**	**-5.96638**	**35.978056**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2BL1	**2black.lugugu**	**white**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2BL2	**2black.lugugu**	**white**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2BL3*	2black.lugugu	white	S.P	-5.96722	35.978611	Hombolo	Dodoma	TZA	22.5	750
H2BL4	**2black.lugugu**	**white**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2W1	**2wild**	**green**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2W2	**2wild**	**black**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2W3*	2wild	black	S.P	-5.96722	35.978611	Hombolo	Dodoma	TZA	22.5	750
H2W4	**2wild**	**green**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2W5	**2wild**	**green**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2W6	**2wild**	**black**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2WL1	**2white.lugugu**	**white**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2WL2*	2white.lugugu	white	S.P	-5.96722	35.978611	Hombolo	Dodoma	TZA	22.5	750
H2WL3	**2white.lugugu**	**white**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2WL4	**2white.lugugu**	**white**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2WL5	**2white.lugugu**	**white**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H2WL6	**2white.lugugu**	**white**	**S.P**	**-5.96722**	**35.978611**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H3LI1	**3limondigua**	**bl.white**	**S.P**	**-5.97222**	**35.984444**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H3MG2	**3magaje**	**orange**	**S.P**	**-5.97222**	**35.984444**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H3N1	**3namata**	**white**	**S.P**	**-5.97222**	**35.984444**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**
H3N2	**3namata**	**white**	**S.P**	**-5.97222**	**35.984444**	**Hombolo**	**Dodoma**	**TZA**	**22.5**	**750**

Acc.	Name	grain.col	Source	Lat	Long	Coll. site	Province	Country	Temp	Prec
H3N3	3namata	white	S.P	-5.97222	35.984444	Hombolo	Dodoma	TZA	22.5	750
H3N4	3namata	white	S.P	-5.97222	35.984444	Hombolo	Dodoma	TZA	22.5	750
H3R1	3roma	white	S.P	-5.97222	35.984444	Hombolo	Dodoma	TZA	22.5	750
H3R2	3roma	white	S.P	-5.97222	35.984444	Hombolo	Dodoma	TZA	22.5	750
H3R3	3roma	white	S.P	-5.97222	35.984444	Hombolo	Dodoma	TZA	22.5	750
H3W1	3wild	green	S.P	-5.97222	35.984444	Hombolo	Dodoma	TZA	22.5	750
H3W2*	3wild	green	S.P	-5.97222	35.984444	Hombolo	Dodoma	TZA	22.5	750
H3W3	3wild	green	S.P	-5.97222	35.984444	Hombolo	Dodoma	TZA	22.5	750
H3W4	3wild	green	S.P	-5.97222	35.984444	Hombolo	Dodoma	TZA	22.5	750
H3W5	3wild	green	S.P	-5.97222	35.984444	Hombolo	Dodoma	TZA	22.5	750
H4BL1	4black.lugugu	white	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4BL2	4black.lugugu	white	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4BL3	4black.lugugu	white	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4BL4	4black.lugugu	white	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4BL5	4black.lugugu	white	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4MG1	4magaje	red	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4R2	4roma	cream	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4R3	4roma	cream	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4W1	4wild	black	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4W2*	4wild	black	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4W3	4wild	green	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4W4	4wild	black	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4W5	4wild	green	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4WL1	4white.lugugu	white	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4WL2	4white.lugugu	white	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4WL3	4white.lugugu	white	S.P	-5.97972	35.980556	Hombolo	Dodoma	TZA	22.5	750
H4WL4	4white.lugugu	white	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H4WL5	4white.lugugu	white	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5BL1	5black.lugugu	white	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5BL2	5black.lugugu	white	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5BL3	5black.lugugu	white	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5BL4	5black.lugugu	white	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5BL5	5black.lugugu	white	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5MG	5magaje	red	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5R1	5roma	cream	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5R2	5roma	cream	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5W1	5wild	orange	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5W2	5wild	green	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5W3	5wild	green	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5W4	5wild	green	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5WL1	5white.lugugu	white	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5WL2	5white.lugugu	white	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5WL3	5white.lugugu	white	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
H5WL4	5white.lugugu	white	S.P	-5.97527	35.981944	Hombolo	Dodoma	TZA	22.5	750
WP1	wild.populatio	black	S.P	-5.98	35.982778	Hombolo	Dodoma	TZA	22.5	750
WP2	wild.populatio	black	S.P	-5.98	35.982778	Hombolo	Dodoma	TZA	22.5	750
WP3*	wild.pop	black	S.P	-5.98	35.982778	Hombolo	Dodoma	TZA	22.5	750
WP4	wild.pop	green	S.P	-5.98	35.982778	Hombolo	Dodoma	TZA	22.5	750
WP5	wild.pop	black	S.P	-5.98	35.982778	Hombolo	Dodoma	TZA	22.5	750

Acc.	Name	grain.col	Source	Lat	Long	Coll. site	Province	Country	Temp	Prec
WP6	wild.pop	black	S.P	-5.98	35.982778	Hombolo	Dodoma	TZA	22.5	750
ML1	Coast	white	NPGRC	NA	NA	NA	NA	TZA	NA	NA
ML10	Coast	white	NPGRC	NA	NA	NA	NA	TZA	NA	NA
ML13	Coast	white	NPGRC	NA	NA	NA	NA	TZA	NA	NA
ML24	Coast	white	NPGRC	NA	NA	NA	NA	TZA	NA	NA
TZA76	Rukwa	White	NPGRC	-7.966667	31.616667	Sumbawanga	Rukwa	TZA	19.3	939
TZA77	Rukwa	white	NPGRC	-7.966667	31.616667	Sumbawanga	Rukwa	TZA	19.3	939
TZA122	Rukwa	yellow	NPGRC	-7.966667	31.616667	Sumbawanga	Rukwa	TZA	19.3	939
TZA123	Rukwa	red	NPGRC	-7.966667	31.616667	Sumbawanga	Rukwa	TZA	19.3	939
TZA124	Rukwa	white	NPGRC	-7.966667	31.616667	Sumbawanga	Rukwa	TZA	19.3	939
TZA189	Mbeya	red	NPGRC	-9.109570	32.936451	Vwawa	Rukwa	TZA	19.9	1223
TZA239	Mbeya	white	NPGRC	-8.417824	33.168289	Makongolosi	Mbeya	TZA	21.9	1133
TZA501	Dodoma	white	NPGRC	-6.173056	35.741944	NA	Mbeya	TZA	22.3	560
TZA517	Dodoma	white	NPGRC	-6.173056	35.741944	NA	Dodoma	TZA	22.3	560
TZA519	Dodoma	white	NPGRC	-5.4833	35.6333	NA	Dodoma	TZA	21.5	634
TZA546	Singida	white	NPGRC	-6.352157	35.214768	NA	Dodoma	TZA	23.4	564
TZA550	Singida	red	NPGRC	-4.5	34.743549	NA	Singida	TZA	20.4	683
TZA560	Singida	white	NPGRC	-4.817443	34.743549	NA	Singida	TZA	20.4	683
TZA607	Singida	yellow	NPGRC	-4.3	34.7	NA	Singida	TZA	20.3	737
TZA612	Singida	white	NPGRC	-4.2	34.4333	NA	Singida	TZA	20	811
TZA625	Mara	orange	NPGRC	-1.775354	34.153195	NA	Singida	TZA	21.9	1039
TZA630	Mara	orange	NPGRC	-1.775354	34.153195	NA	Mara	TZA	21.9	1039
TZA675	Mara	red	NPGRC	-1.775354	34.153195	NA	Mara	TZA	21.9	1039
TZA691	Mara	red	NPGRC	-1.775354	34.153195	NA	Mara	TZA	21.9	1039
TZA1568	Kagera	red	NPGRC	-1.3	31.416667	Kilimile	Mara	TZA	20.8	1345
TZA1587	Kagera	red	NPGRC	-1.1	30.816667	Bugomola	Kagera	TZA	19.3	979
TZA4146	Kagera	red	NPGRC	-2.905	31.198333	Biharamulo	Kagera	TZA	20.4	966
TZA4162	Kagera	white	NPGRC	-3.106667	31.136389	Biharamulo	Kagera	TZA	21.4	948
TZA2265	Mtwara	white	NPGRC	-10.002080	-39.706959	Mahumbika	Kagera	TZA	26	924
TZA2328	Mtwara	white	NPGRC	-10.0833	39.85	Nachunyu	Mtwara	TZA	26.2	935
TZA2331	Mtwara	white	NPGRC	-10.0833	39.85	Nachunyu	Mtwara	TZA	26.2	935
TZA2347	Mtwara	white	NPGRC	-10.594130	39.666229	Njengwa	Mtwara	TZA	24.3	1105
TZA2357	Mtwara	white	NPGRC	-10.594130	39.666229	Mruma	Mtwara	TZA	24.3	1105
TZA2675	Morogoro	white	NPGRC	-6.167778	37.667222	Turiani	Morogo	TZA	25.1	1030
TZA2717	Morogoro	white	NPGRC	-7.064167	36.9025	Morogoro	Morogo	TZA	24.1	963
TZA2726	Morogoro	white	NPGRC	-7.199722	36.926944	Morogoro	Morogo	TZA	23.9	1014
TZA3138	Kigoma	white	NPGRC	-4.728333	29.866944	Munzeze	Morogo	TZA	22.5	1085
TZA3147	Kigoma	red	NPGRC	-4.624444	30.258889	Kasulu	Kigoma	TZA	21.2	1114
TZA3228	Kigoma	white	NPGRC	-3.618056	30.4675	Kibondo	Kigoma	TZA	21.3	1163
TZA3238	Tabora	white	NPGRC	-4.978611	32.386389	Ufuluma	Tabora	TZA	23.4	985
TZA4004	Mwanza	red	NPGRC	-2.092222	33.027222	Nansio	Mwanza	TZA	22.4	1353
TZA4009	Mwanza	red	NPGRC	-2.116944	33.151111	Nansio	Mwanza	TZA	22.6	1286
TZA4011	Mwanza	red	NPGRC	-2.111111	33.044167	Nansio	Mwanza	TZA	22.5	1337
IS 3569	Sudan	yellow	ICRISAT	4.2399998	32.330002	Torit	NA	SDN	25	1190
IS 3544	Sudan	NA	ICRISAT	4.54	29.27	Maridi	NA	SDN	24.5	1509
IS 22380	Sudan	NA	ICRISAT	13.83	35.43	Kassab	NA	SDN	28.4	628
IS 2319	Sudan	bl. white	ICRISAT	9.54	32.07	Kadok	NA	SDN	28	805
IS 26834	Sudan	pale red	ICRISAT	4.6300001	32.630001	Loronjo	NA	SDN	27.2	887
IS 26836	Sudan	bl. white	ICRISAT	4.6300001	32.630001	Loronjo	NA	SDN	27.2	887

Acc.	Name	grain.col	Source	Lat	Long	Coll. site	Province	Country	Temp	Prec
IS 26841	Sudan	pale red	ICRISAT	4.79	32.689999	Mura-ikotos	NA	SDN	23.5	1089
IS 7305	Nigeria	bl.white	ICRISAT	11.56	13.01	Masba	NA	NGA	25.6	682
IS 710	Nigeria	orange	ICRISAT	11.48	11.1	Babana	NA	NGA	24.6	757
IS 7957	Nigeria	orange	ICRISAT	11.4	4.11	Rima	NA	NGA	27.9	951
IS 9108	Kenya	red	ICRISAT	-0.07	34.810001	Kibos	NA	KEN	22.6	1306
IS 9113	Kenya	red	ICRISAT	-0.07	34.810001	Kibos	NA	KEN	22.6	1306
IS 21512	Malawi	white	ICRISAT	-16.15	34.779999	Kashonte	NA	MWI	25.9	783
IS 2205	India	white	ICRISAT	28.625514	77.181702	NA	NA	IND	25	698
IS 22294	Botswana	orange	ICRISAT	-23.01	27.76	Seleka	NA	BWA	21.2	380
IS 22720	Somalia	cream	ICRISAT	3.4200001	43.5	Bulo-burlow	NA	SOM	26.9	495
IS 22799	Somalia	red	ICRISAT	4.75	45.25	Look-Jelow	NA	SOM	28.6	266
IS 11473	Ethiopia	bl.white	ICRISAT	15.1	36.650002	Teseney	NA	ETH	28.4	390
IS 11619	Ethiopia	red	ICRISAT	9.0299997	38.700001	Addis Ababa	NA	ETH	15.4	1166
IS 23586	Ethiopia	white	ICRISAT	8.2700005	34.66	Fumaro	NA	ETH	27.2	1163
IS 23590	Ethiopia	white	ICRISAT	8.2700005	34.66	Fumaro	NA	ETH	27.2	1163
IS 25732	Mali	white	ICRISAT	14.27	-10.36	Diakautame	NA	MLI	27.5	789
IS 25989	Mali	red	ICRISAT	11.28	-7.01	Tienaga	NA	MLI	26.9	1041
IS 26025	Mali	white	ICRISAT	12.39	-6.36	Nangola	NA	MLI	27.3	926
IS 2379	South Africa	bl. white	ICRISAT	-23.83333	31.5	Transvaal	NA	ZAF	22.7	505
IS 2382	South Africa	red	ICRISAT	-23.83333	31.5	Transvaal	NA	ZAF	22.7	505
IS 27887	South Africa	bl. white	ICRISAT	-24.19	24.9	Skuinsrand	NA	ZAF	17.9	621
IS 27912	South Africa	pale red	ICRISAT	-26	29.76	Groenland	NA	ZAF	14.8	729
IS 29358	Lesotho	white	ICRISAT	-29.51	27.620001	Maseru	NA	LSO	14.7	713
IS 29392	Lesotho	white	ICRISAT	-29.5	28	Maseru	NA	LSO	9.3	901
IS 29441	Lesotho	bl.white	ICRISAT	-30.13	28.700001	Qachas nek	NA	LSO	13.9	723
IS 29468	Lesotho	white	ICRISAT	-30.02	27.549999	Mpharane	NA	LSO	13.9	715
IS 29689	Zimbabwe	white	ICRISAT	-19.52	31.629999	Canary	NA	ZWE	19.7	580
IS 29772	Zimbabwe	red	ICRISAT	-18.53	32.119999	Rusape	NA	ZWE	17.2	796
IS 29733	Zimbabwe	white	ICRISAT	-20.33	30.040001	Zvishavane	NA	ZWE	20.1	570
IS 29914	Zimbabwe	orange	ICRISAT	-20.03	29.1	Inyozan	NA	ZWE	18.7	625
IS 31043	Uganda	pale red	ICRISAT	3.3499999	33.330002	Naam Okara	NA	UGA	23.7	998
IS 31186	Uganda	orange	ICRISAT	1.8	33.549999	Ebijiing	NA	UGA	24.2	1336
IS 8916	Uganda	red	ICRISAT	0.46	34.080002	Busia	NA	UGA	22.1	1618
IS 32831	Tanzania	red	ICRISAT	-6.28	36.75	Mlali	Dodoma	TZA	20.8	742
IS 32862	Tanzania	red	ICRISAT	-4.9	35.779999	Ghadua	Singida	TZA	21.1	719

Table Apx2 Information on the subdivision of the complete dataset into six smaller datasets, the definition of groups within these and the analysis performed on each dataset. The complete dataset, consisted of 161 sorghum accessions, included accessions from three geographical scales; 1) Africa- consists of 41 cultivated sorghum accessions from throughout Africa, 2) Tanzania- consists of 42 cultivated sorghum accessions from throughout Tanzania and 3) Hombolo- consists of 52 cultivated sorghum accessions (landraces) and 26 'wild' accessions ('wild') collected from five households in Hombolo, Tanzania.

Datasets	Analysis performed and definition of groups for analysis
Complete dataset (Africa, Tanzania and Hombolo; 161 accessions)	STRUCTURE analysis, and calculations of Rs, PRs. For calculations of Rs and PRs, the dataset was divided into three groups, each group consisting of the accessions belonging to Hombolo, Tanzania and Africa, respectively.
Africa (41 accessions)	STRUCTURE analysis, Mantel test, and AMOVA. For AMOVA groups were defined based on 1) STRUCTURE analysis (four groups), 2) geography (11 groups) grain color (two groups), 3) grain color (two groups), 4) race (seven groups), 5) precipitation (eight groups) and 6) temperature (eight groups). For the AMOVA (geography) accessions from Botswana, Malawi, and India were excluded because there was only one accession from these countries.
Tanzania (42 accessions)	STRUCTURE analysis, Mantel test, AMOVA and PCA. For AMOVA groups were defined based on 1) STRUCTURE analysis (five groups), 2) geography (11 groups), 3) grain color (two groups), 4) precipitation (seven groups) and 5) temperature (seven groups). For the PCA and AMOVA (geography) 11 groups were defined based on province. The province 'Tabora' was excluded because there was only one accession from this province.
Hombolo (landraces; 52 accessions)	STRUCTURE analysis, PCO, Mantel test, linkage disequilibrium, and AMOVA. For AMOVA, groups were defined based on 1) landraces (five groups), 2) households (five groups), 3) grain color (two groups) and 4) STRUCTURE analysis (eight groups). For calculations of He, Nm, PRs and Rs groups were defined based on landraces (five groups). For pairwise Rst calculations groups were defined based on landraces (eight groups). For calculations of He, Nm, PRs, Rs and AMOVA at least two accessions per group were needed. For the definition of groups based on landraces, the landraces wawa, limondigua and sandala (each with only one accession) were excluded.
Hombolo (landraces and 'wild'; 78 accessions)	STRUCTURE analysis, PCO, PCA, Mantel test, calculations of He, Rs, PRs, and pairwise Rst. For AMOVA and calculations of Nm, He, Rs and PRs two groups were defined based on cultivated versus 'wild' accessions. For pairwise Rst six groups were defined; The first five groups consisted of accessions belonging to household (1-5) and the last groups consisted of accessions from the 'wild' population.
Hombolo ('wild';26 accessions)	Mantel test, and calculations of linkage disequilibrium.
Hombolo (landraces) and Tanzania (94 accessions)	STRUCTURE analysis, PCO and neighbor joining analysis.

Table Apx3 AMOVA for cultivated and 'wild' sorghum based on 17 microsatellite markers. Africa- refers to 41 cultivated accessions from throughout Africa. Tanzania- refers 42 cultivated accessions from throughout Tanzania, and Hombolo- refers to 26 'wild' and 52 cultivated accessions collected from five households in Hombolo, Tanzania. Groups were defined for Africa based on 1) STRUCTURE analysis (K=4), 2) geography, 3) grain color (red or white), 4) race, 5) mean temperature at the collection sites (temp) and 6) mean annual precipitation from the collection sites (prec). Groups were defined for the accessions from Tanzania based on 1) STRUCTURE analysis (K=5), 2) geography, 3) grain color (red or white), 4) mean temperature of the collection sites (temp) and 5) mean annual precipitation at the collection sites (prec). Groups were defined for cultivated accessions from Hombolo based on 1) STRUCTURE analysis (K=8), 2) household affiliation, 3) landraces, 4) grain color (red or white) and 5) cultivated (cult) versus wild sorghum (wild). S.sq.= sum of squares, V.comp.=variance components, Perc.var.=percentage variations. Significance tests consisted of 1640 permutation.

Source of variation	S.sq.	V. comp.	Perc.var.	Source of variation	S.sq.	V.comp.	Perc.var.
Africa (STRUCTURE)				**Tanzania (grain color)**			
Among populations	95.49	1.07	16.58	Among populations	33.80	0.65	10.71
Within populations	385.09	5.29	81.71	Within populations	369.00	4.46	73.61
Within accessions	4.50	0.11	1.71	Within accessions	37.00	0.95	15.66
Total	485	6.47		Total	440.00	6.00	
Africa (geography)				**Tanzania (temp)**			
Among populations	158.08	0.72	11.82	Among populations	82.95	0.38	6.69
Within populations	285.00	5.42	86.25	Within populations	292.00	4.52	77.95
Within accessions	4.50	1.12	1.91	Within accessions	32.00	0.89	15.00
Total	448.00	6.28		Total	407.45	5.79	
Africa (grain color)				**Tanzania (prec)**			
Among populations	34.07	-0.006	-1.01	Among populations	76.31	0.28	4.88
Within populations	457.71	6.16	99.26	Within populations	281.31	4.61	79.08
Within accessions	4.50	0.10	1.74	Within accessions	32.00	0.95	16.00
Total	496.29	6.02		Total	390.00	5.83	
Africa (race)				**Hombolo (STRUCTURE)**			
Among populations	93.32	0.47	7.39	Among populations	161.00	1.58	33.85
Within populations	273.00	5.81	91.07	Within populations	193.00	1.77	37.81
Within accessions	3.00	0.09	1.53	Within accessions	63.00	1.32	28.32
Total	702	4.79		Total	419.00	4.69	
Africa (temp)				**Hombolo (households)**			
Among populations	106.87	0.52	8.22	Among populations	81.43	0.69	15.07
Within populations	247.61	5.72	89.92	Within populations	295.35	2.68	58.49
Within accessions	3.50	0.11	1.85	Within accessions	60.00	1.21	26.42
Total	357.29	6.36		Total	437.29	4.58	
Africa (prec)				**Hombolo (landraces)**			
Among populations	76.87	0.20	4.77	Among populations	72.18	0.67	14.72
Within populations	282.91	4.60	80.47	Within populations	280.39	2.54	55.52
Within accessions	33.00	0.94	15.00	Within accessions	66.00	1.36	29.75
Total	394.29	5.22		Total	418.57	4.58	
Tanzania (STRUCTURE)				**Hombolo (grain color)**			
Among populations	122.60	1.49	24.55	Among populations	22.29	1.04	19.33
Within populations	281.23	3.61	59.57	Within populations	365.50	3.05	56.51
Within accessions	38.00	0.96	15.86	Within accessions	67.00	1.37	24.15
Total	441.00	6.00		Total	454.79	5.41	
Tanzania (geography)				**Hombolo (cult, wild)**			
Among populations	162	1.11	19.12	Among populations	16.74	0.14	3.13
Within populations	235	3.79	64.95	Within populations	587.00	3.29	69.66
Within accessions	36	0.93	15.94	Within accessions	98.00	1.28	27.24
Total	433	5.84		Total	702.00	4.79	

Table Apx4 **Landraces of sorghum sampled from five households in Hombolo, Dodoma, Tanzania, 3-5 June 2011. From each plant both grains and leaf material were collected.**

Landraces	Household1	Household2	Household3	Household4	Household5
Black lugugu	5 plant	5 plants	-	5 plants	5 plants
White lugugu	5 plant	5 plants	-	5 plants	5 plants
Namata	-	-	5 plants	-	-
Roma	-	-	1 plant	1 plant	1 plant
Magaje	-	-	1 plant	1 plant	1 plant
Wawa	1 plant	-	-	-	-
Limondigua	-	-	1 plant	-	-
Sandala	1 plant	-	-	-	-

Table Apx5 **Questionnaire interviews with five farmers, regarding their sorghum crop, in Hombolo, Dodoma, Tanzania, 3-5 June 2011.**

Question	Household 1	Household 2	Household 3	Household 4	Household 5
Name?	Hegla Manjono	Hegla Sazina	Rebecca Manyono	Unis and Stefano Massima	Paulo Chiluka
Are you mixing landraces in your field?	Yes	Yes- also with pearl millet- Sometimes white and black lugugu are separated	Yes	Yes	Yes
Where did you get your seeds from?	Neighbors	-	Neighbors	Neighbors	-
Have you recycled your grains?	Recycled more than 10 years	-	Recycled for four years	Recycled some years	Recycled many years
Do you find wild sorghum? Are they a problem?	Yes. They are a problem	Yes. They are a problem	Yes. They are a problem	Yes. They are a problem	Yes. They are a problem
Why use local landraces?	Storage and palatability	Storage and palatability	Storage and palatability	Storage and palatability	Storage and palatability

Table Apx6 **Genotypic linkage disequilibrium based on 17 microsatellite markers for 26 'wild' sorghum accessions and 52 cultivated sorghum accessions, collected from five households in Hombolo, Tanzania. Only the locus pairs with p ≤0.05 are shown.**

'wild' sorghum			Cultivated sorghum		
Locus#1	**Locus#2**	**P-Value**	**Locus#1**	**Locus#2**	**P-Value**
Xtxp40	sb5-236	0.01561	Xcup02	Xtxp123	0.043950
sb5-236	Xtxp57	0.003150	Xcup02	mSbCIR283	0.003830
sb5-236	mSbCIR283	0.002250	sb5-236	Xtxp295	0.001900
Xcup02	mSbCIR283	0.025330	Xcup02	Xtxp295	0.001450
sb5-236	Xtxp295	0.000000	sb5-236	Xcup61	0.000350
mSbCIR283	Xtxp295	0.000000	sb5-236	Xtxp289	0.012080
Xtxp57	Xcup61	0.013500	Xcup02	Xtxp289	0.014080
mSbCIR283	Xcup61	0.000250	Xtxp57	Xtxp289	0.006700
Xtxp295	Xcup61	0.000000	Xcup14	Xtxp289	0.000000
sb5-236	Xcup14	0.031350	sb5-236	Xgap206	0.009110
Xtxp295	Xcup14	0.005680	Xtxp57	Xgap206	0.001100
sb5-236	Xtxp289	0.000000	mSbCIR283	Xgap206	0.046530
Xtxp57	Xtxp289	0.000150	Xcup14	Xgap206	0.000000
mSbCIR283	Xtxp289	0.017570	Xtxp289	Xgap206	0.029840
Xtxp295	Xtxp289	0.044500	sb5-236	Xtxp320	0.007080
Xcup61	Xtxp289	0.039210	Xcup02	Xtxp320	0.007770
Xcup14	Xtxp289	0.009550	Xtxp57	Xtxp320	0.000000
sb5-236	Xgap206	0.000000	mSbCIR283	Xtxp320	0.008480
mSbCIR283	Xgap206	0.000510	Xtxp295	Xtxp320	0.029530
Xtxp289	Xgap206	0.005090	Xcup14	Xtxp320	0.002620
sb5-236	Xtxp320	0.010600	Xtxp289	Xtxp320	0.000000
Xcup14	Xtxp320	0.000000	Xgap206	Xtxp320	0.000000
Xtxp289	Xtxp320	0.027560	sb5-236	sbAG02	0.017820
Xtxp40	XtxpXcup141	0.016930	Xcup02	sbAG02	0.011170
mSbCIR283	XtxpXcup141	0.001460	Xtxp57	sbAG02	0.000720
Xgap206	XtxpXcup141	0.042490	Xcup61	sbAG02	0.001700
sbAG02	XtxpXcup141	0.019730	Xtxp289	sbAG02	0.007800
sb5-236	Xtxp12	0.028920	Xtxp320	sbAG02	0.018160
mSbCIR283	Xtxp12	0.000000	sb5-236	Xtxp141	0.011290
Xtxp295	Xtxp12	0.025030	Xcup02	Xtxp141	0.005150
Xcup61	Xtxp12	0.008900	mSbCIR283	Xtxp141	0.000000
Xtxp289	Xtxp12	0.002990	Xcup61	Xtxp141	0.006040
Xgap206	Xtxp12	0.000000	Xtxp289	Xtxp141	0.000000
XtxpXcup141	Xtxp12	0.013980	sbAG02	Xtxp141	0.000000
sb5-236	Xtxp15	0.000000	mSbCIR283	Xtxp12	0.000000
Xtxp57	Xtxp15	0.047470	Xtxp295	Xtxp12	0.000000
Xtxp295	Xtxp15	0.006730	Xcup61	Xtxp12	0.002880
Xcup61	Xtxp15	0.038620	Xcup14	Xtxp12	0.003740
Xcup14	Xtxp15	0.012490	Xtxp289	Xtxp12	0.000000
Xtxp289	Xtxp15	0.000270	Xgap206	Xtxp12	0.035380
Xgap206	Xtxp15	0.012940	Xtxp141	Xtxp12	0.012960
Xtxp320	Xtxp15	0.009070	sb5-236	Xtxp15	0.000280
XtxpXcup141	Xtxp15	0.021700	Xtxp57	Xtxp15	0.031060
Xcup61	Xtxp278	0.023510	mSbCIR283	Xtxp15	0.000000
Xgap206	Xtxp278	0.024490	Xcup61	Xtxp15	0.005900
			Xcup14	Xtxp15	0.040900
			Xtxp289	Xtxp15	0.003520
			Xgap206	Xtxp15	0.002840
			sbAG02	Xtxp15	0.004710
			Xtxp141	Xtxp15	0.000950
			Xtxp12	Xtxp15	0.000000

Table Apx7 Significance values from a Wilcoxon Signed Rank test, for allelic richness (Rs), private allelic richness (PRs) and expected heterozygosity (He), across 17 microsatellites, for 52 cultivated and 26 'wild' sorghum accessions collected from five households in Hombolo, Tanzania. The test compares different means

with the null hypothesis "the means are the same". If the p value is below 0.05 then there is 95% certainty that the means are different. BL=black lugugu, WL=white lugugu, N=namata, R=roma, Mg=magaje, cult=all landraces, wild='wild' sorghum.

P values Rs	P values PRs	P values He
p(BL,WL)=0.190	p(BL,WL)=0.500	p(BL,WL)=0.670
p(WL,N)=0.001	p(WL,N)=0.400	p(WL,N)=0.010
p(BL,N)=0.001	p(BL,N)=0.500	p(BL,N)=0.004
p(R,Mg)=0.300	p(R,MG)=0.100	p(R,MG)=0.420
p(WL, R)=0.400	p(WL,R)=0.200	p(WL,R)=0.350
p(WL,Mg)=0.130	p(WL,Mg)=0.300	p(WL,MG)=0.170
p(BL,Mg)=0.030	p(BL,Mg)=0.200	p(BL,MG)=0.050
p(BL,R)=0.080	p(BL,R)=0.300	p(BL,R)=0.055
p(N,R)=0.030	p(N,R)=0.700	p(N,R)=0.140
p(Mg,N)=0.500	p(MG,N)=0.500	p(MG,N)=0.770
p(cult,wild)=0.630	p(cult,wild)=0.640	p(cult,wild)=0.780

Figures

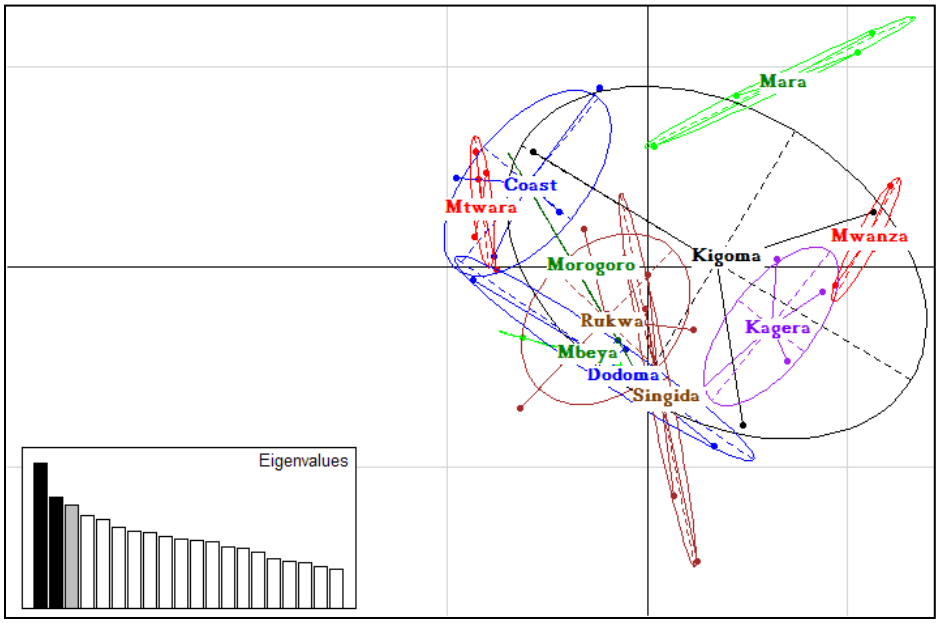


Figure Apx1 **Plot of the first and second components of a Principal Component Analysis (PCA) based on 17 microsatellite markers, for 42 sorghum accessions from 11 provinces in Tanzania (indicated with different colors). Eigenvalues corresponding to the two components are filled in black. Each point represents a sorghum accession, and is connected to the mean point of its group by a line of similar color. The ellipses show the 95% confidence limit around the mean of the group.**

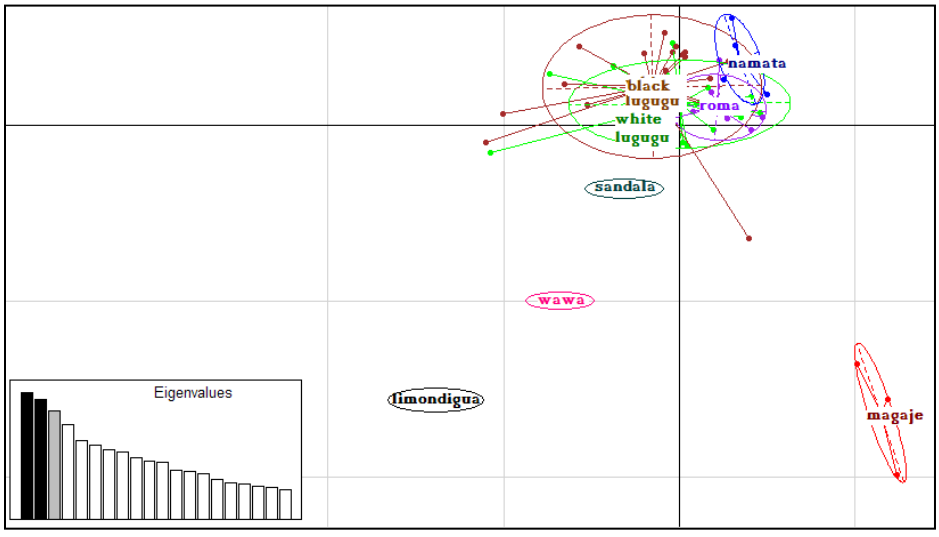


Figure Apx2 **Plot of the first and second components of a Principal Component Analysis (PCA) based on 17 microsatellite loci, for 52 accessions of cultivated sorghum, representing seven landraces (indicated with different colors) sampled from five households in Hombolo, Tanzania. Eigenvalues corresponding to the two components are filled in black. Each point represents a sorghum accession, and is connected to the mean point of its group by a line of similar color. The ellipses show the 95% confidence limit around the mean of the group.**

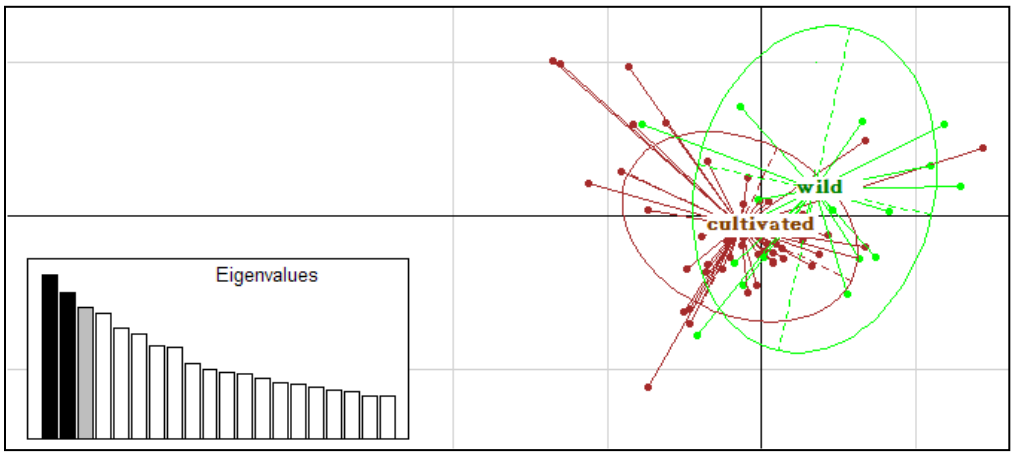


Figure Apx3 **Plot of the first and second components of a Principal Component Analysis (PCA) based on 17 microsatellite loci, for 52 cultivated and 26 'wild' sorghum accessions (indicated with different colors) sampled from five households in Hombolo, Tanzania. Eigenvalues corresponding to the two components are filled in black. Each point represents a sorghum accession, and is connected to the mean point of its group by a line of similar color. The ellipses show the 95% confidence limit around the mean of the group.**

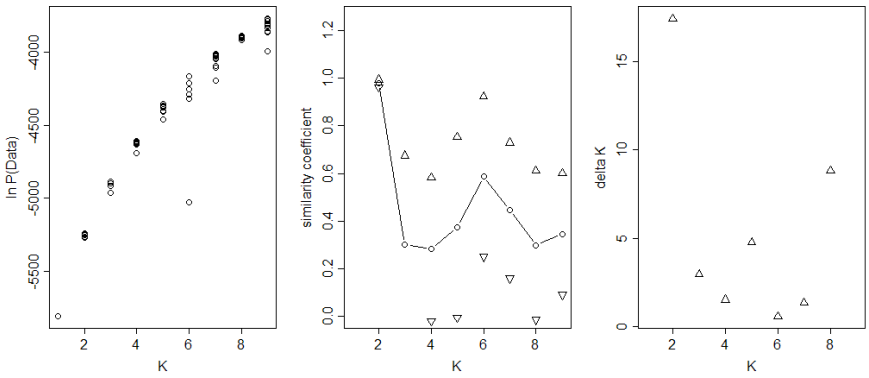


Figure Apx4 **Plot of the output from STRUCTURE-SUM for 94 cultivated sorghum accessions, based on 17 microsatellites. The plot includes a summary, for K= 1-9, of the logarithmic probability (Ln P(D)), the similarity coefficient and delta K.**

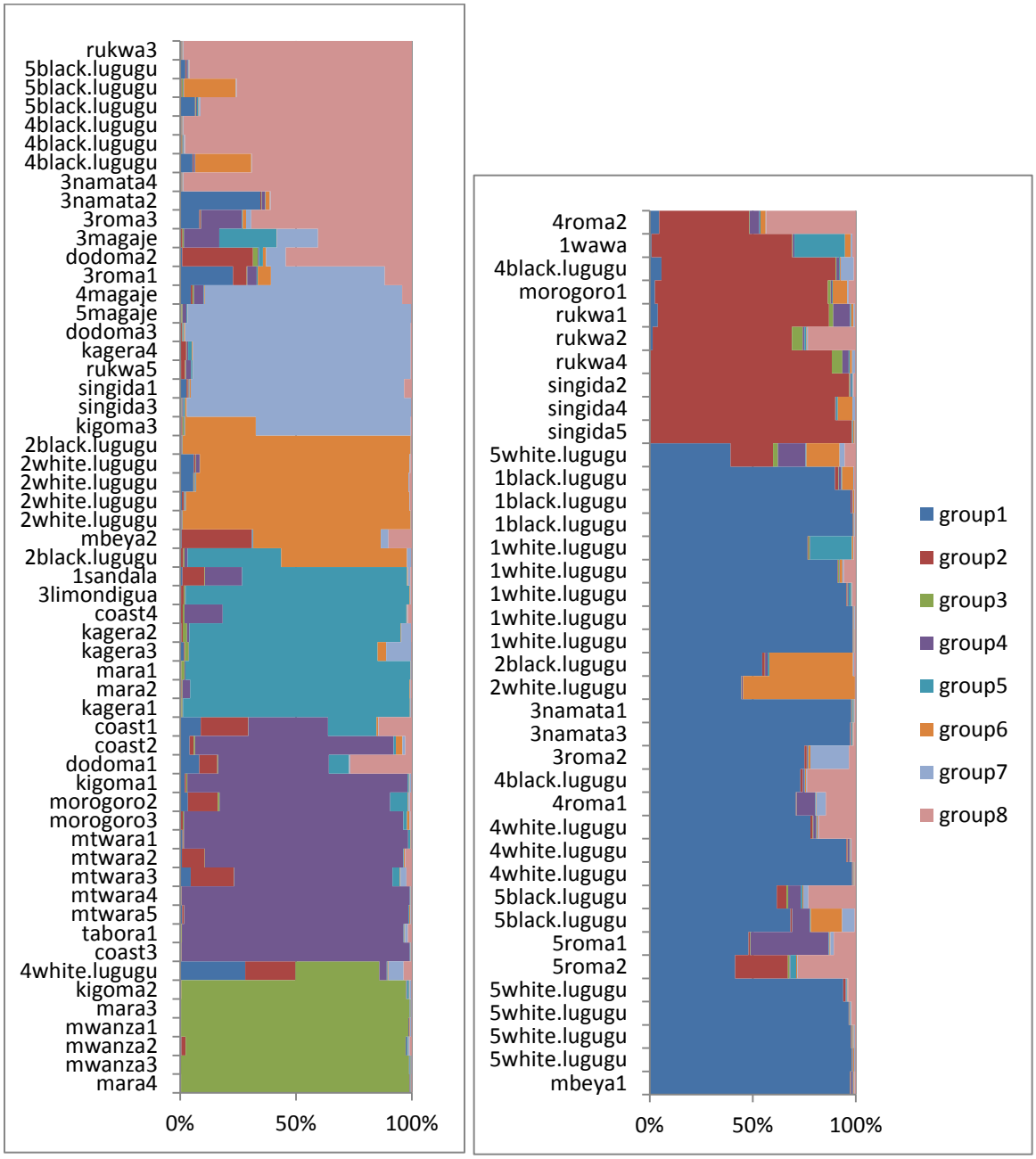


Figure Apx5 **Plot of the STRUCTURE results (K=8) for cultivated sorghum, based on 17 microsatellites. The accessions represent two geographical scales 1) local scale- 52 sorghum accessions, collected from five households in Hombolo, Tanzania, representing eight landraces (white lugugu, black lugugu, namata, roma, magaje, limondigua, sandal and wawa) the number preceding the landrace names refers to the household from which they were collected and 2) country scale- 42 sorghum accessions from throughout Tanzania. The STRUCTURE groups (K) are represented by different colors. The segmentation of the horizontal pillars shows with what percentage an accessions is placed within which groups.**